Structural characterisation of Histidine Kinase 2

Liang Wang

School of Biological and Chemical Sciences

Queen Mary University of London

Supervisors: Richard Pickersgill and Jon Nield

Submitted in partial fulfilment of the requirements of the Degree of

Doctor of Philosophy

Statement of originality

This thesis includes ideas originally developed by Prof. John F. Allen, who is now at University College London, and incorporates results of experiments undertaken in collaboration with Dr. Iskander Ibrahim in the laboratory of Prof. John F. Allen at Queen Mary University of London, U.K., and with Ms. Liyana Azmi, under the supervision of Prof. Olwyn Byron at University of Glasgow, U.K. These results are pointed out in the text in chapter 3 and concern experiments to help elucidate the oligomeric states of Histidine Kinase 2 (Hik2). This thesis also incorporates results of single particle analysis processing by Dr. Jon Nield at Queen Mary University of London, U.K., all as indicated in chapter 5.

I certify that, with the above qualifications, this thesis, and the research it describes, are entirely the product of my own work. Any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices.

Liang Wang

May 2017

Abstract

Two-component systems (TCS) are the predominant signal transduction pathways in prokaryotes, being present also in eukaryotic organisms, such as algae, fungi and yeast, and higher plants. TCSs play an important role in environmental signal perception and response, essentially implementing adaptation to the surrounding environment. Histidine Kinase 2 (Hik2) in cyanobacteria is a typical sensor histidine kinase, one component of a TCS, and has been identified to be a homologue protein of *Arabidopsis* Chloroplast Sensor Kinase (CSK). Previous research has elucidated Hik2 to regulate photosynthetic gene transcription with two response regulators, Rre1 and RppA via phosphorylation.

A typical histidine kinase contains a variable sensor domain and a conserved kinase domain. It usually functions as a homodimer. This thesis describes the structural characterisation of Hik2, probing particularly its discovered oligomeric states. Results obtained from size exclusion chromatography, native-PAGE, chemical cross-linking analyses and mass spectrometry, amongst others, have shown a variety of Hik2 structural populations exist, further validated by negative stain transmission electron microscopy coupled to single particle analysis. Hik2 protein exists predominantly as a hexamer in low salt conditions, and adding NaCl dissociates hexamers into tetramers, critical for the autophosphorylation activity of Hik2. Thus, a model is proposed for the constitution change of Hik2 oligomers when salt concentration differs. In addition, the sensor domain is typically responsible for detecting environmental input, however, it is not yet clear how Hik2 and CSK sense signals. In this thesis, the structures of Hik2 and CSK sensor domains were analysed and discussed, to aid our understanding of their mechanism of signal perception and transduction.

Acknowledgements

My Ph.D. studies have been somewhat bumpier than I expected, but I am very grateful and extremely lucky in getting support from so many nice people. I would like to begin these acknowledgements by sincerely thanking my supervisors. I would like to express my utmost gratitude to Prof. Richard W. Pickersgill for his supervision in my later Ph.D. stages. He took me in as his Ph.D. student and gave me support when I needed it most, and gave me freedom to do my own research. Without his guidance and encouragement, my studies would have been much less fruitful.

I want to sincerely thank my first supervisor Prof. John F. Allen, who offered this amazing opportunity to begin my Ph.D. at Queen Mary University of London, and always cared for me, even when he did not need to. He is a model to me, a decent person who is always passionate in science. Thank you for sharing this amazing project with me and showing the beauty of science to me.

A big thank you goes to my second supervisor Dr. Norbert Krauβ. When the big change happened and I was lost and worried, he kindly agreed to be my supervisor and took care of me. His support helped me go through this hard time, and gave me courage to continue with my studies. I would not go this far without his help.

Lastly, I wish to thank my co-supervisor Dr. Jon M. Nield, who not only taught me the TEM technique, but has always been constantly supporting me, from the beginning of my Ph.D. to the end. In the later stages of my Ph.D., he provided me the opportunity to work with him two days per week and spent much time and energy to help with my thesis.

I also want to thank all the support staff I have encountered over the five years: my Ph.D. panel members, Dr. Brendan Curran and Dr. Paul Hurd for their advice and support; Dr. Kristina Zubow for running lab 1.28 and her technical advice for many experiments; Yumi, Lingzhi, Giulia, Gianna, and Ruth for their technical support; Sam for organising so many things for moving the lab. I would like to thank my collaborators Prof. Olwyn Byron and Ms. Liyana Azmi at University of Glasgow for their help with analytical ultracentrifugation and small-angle X-ray scattering experiments.

I give my gratitude to the financial support of China Scholarship Council and Queen Mary University of London.

Special thanks must go to all my friends at Queen Mary, past and present, who have helped create a warm environment. I would like to thank Iskander, who taught me all the basic experimental skills with great patience when I firstly came to a new lab in a foreign country, and thanks for his great trust when we have collaborated, and continue to do so. I thank Wilson for encouraging me not to quit. I thank Nana and Wencai for being wonderful friends and bringing me pleasure and amusement. I thank Junyi for being such a nice roommate who made our flat a cosy place to rest and relax in. A big thank you goes to Shuang, who cared for me all the time, not only helped me with my study but also my life; she is like a sister whom I can get comfort from.

I would like to thank my family, especially my Mom, Zhezhen Guo, my Dad, Qimeng Wang, my two Grandmas, Yinping Zhao, Chunfen Li, and my Grandpa, Zhipeng Wang, for helping to make me who I am today; for all your support and encouragements. I thank my husband Shuai for his great patience and endless love to me, and having taken care of my family when I was far away. Finally, I want to thank my son Ziwu, for

recognising me at his first glance after months long separation when he was only one year old, and willing to take me in without a second thought when I finally come back home!

Reading my book, under the flowering tree, 在鲜花盛开的树下读书，
Listless, my shadow soars above my side, 懒洋洋的影子伴我在身旁。
To the songs I sing the sun flickers her beams, 太阳燃放光芒照耀我的歌声，
Long may we share this fine, flowing tide. 愿我们共享流金岁月，久久长长。

With homage to Bai Li.

Table of contents

List of figures

List of tables

Abbreviations

2D	two-dimensional
3D	three-dimensional
Å	Ångstrom
ADP	adenosine 5'-diphosphate
ATP	adenosine 5'-triphosphate
BphP	bacteriophytochrome photoreceptor
CA domain	catalytic and ATP-binding domain
cAMP	cyclic adenosine monophosphate
CCD	charge coupled device
CD	circular dichroism
CSK	chloroplast sensor kinase
CTF	contrast transfer function
CV	column volume
Cyt	cytochrome
DBMIB	dibromothymoquinone
DHp domain	dimerisation and histidine phosphorylation domain
DLS	dynamic light scattering
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
*Dr*BphP	*Deinococcus radiodurans* phytochrome
DSP	dithiobis (succinimidylpropionate)
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
ESI	electrospray ionisation
FAD	flavin adenine dinucleotide
FD	ferredoxin
FFT	fast Fourier transform
FMN	flavin mononucleotide
FNR	ferredoxin $NADP^+$ reductase
FRC	Fourier ring correlation

FSC	Fourier shell correlation
GAF domain	cGMP phosphodiesterase/adenylcyclase/FhlA domain
HAMP domain	Histidine kinases/Adenylyl cyclases/Methyl-accepting chemotaxis proteins/Phosphatase domain
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hik	histidine kinase
Hpt	histidine-containing phosphotransfer domain.
IPTG	isopropyl β-D-1-thiogalactopyranoside
I-TASSER	Iterative Threading ASSEmbly Refinement (prediction server)
LB	lysogeny broth
LBD	ligand binding domain
LHC	light harvesting complex
LOMETS	Local Meta-Threading-Server
LOV domain	light/oxygen/voltage domain
MS	mass spectrometry
$NADP^+$	nicotinamide adenine dinucleotide phosphate oxidised form
NADPH	nicotinamide adenine dinucleotide phosphate
NHS	N-hydroxysuccinimide
NMR	nuclear magnetic resonance
NTP	nucleoside triphosphate
OD	optical density
PAS domain	Per/Arnt/Sim domain
PC	plastocyanin
PCR	polymerase chain reaction
PDB	protein data bank
PDC fold	PhoQ/DcuS/CitA fold
PEP	plastid-encoded RNA polymerase
Pfr-form	the far-red light absorbing form of phytochromes
Phyre2	Protein Homology/analogY Recognition Engine
PMSF	phenylmethylsulphonyl fluoride
PQ	plastoquinone
PQH_2	plastoquinol
Pr-form	the red-light absorbing form of phytochromes

PS I	photosystem I
PS II	photosystem II
PTA	phosphotungstic acid
PTK	plastid transcription kinase
redox	reduction-oxidation
RMSD	root-mean-square deviation
RNA	ribonucleic acid
RppA	regulator of photosynthesis and photopigment-related gene expression A
Rre	response regulator
S	Svedberg
$s_{20,w}$	sedimentation coefficient of a particle in water at 20 ºC
SAXS	small angle X-ray scattering
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SE	sedimentation equilibrium
SIG1	sigma factor-1
SMART	Simple Modular Architecture Research Tool
SOC	super optimal broth with catabolite repression
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
SV	sedimentation velocity
TCS	two-component signal transduction system
TEM	transmission electron microscopy
TEMED	tetramethylethylenediamine
Tm	melting temperature of a primer
TM	transmembrane
TMAO	trimethylamine-N-oxide
TMHMM	TransMembrane prediction using Hidden Markov Models
Tris	tris(hydroxymethyl)aminomethane
UA	uranyl acetate
US-SOMO	UltraScan Solution Modeler
UV	ultraviolet

Chapter 1 **Introduction**

1.1 Two-component signal transduction system

Histidine Kinase 2 (Hik2) investigated in this study has been identified to be a classic Two-Component System (TCS) protein (Ashby and Houmard 2006). TCSs were discovered in prokaryotic bacteria as a predominant method for signal transduction pathways, functioning in cell-cell communication, environmental signal perception and essentially to implement eventual adaptation to the surrounding environment. A typical TCS consists of two conserved proteins, a sensor histidine kinase, and a response regulator (Stock et al. 1985, Nixon et al. 1986). When the organism's environment changes, the sensor histidine kinase perceives the signal, and autophosphorylates at a particular Histidine residue (His), creating a phosphoryl group with high energy to react to the external cue. The phosphoryl group is subsequently transferred to an aspartate residue in the response regulator, which causes a conformational change of its regulatory domain, inducing an appropriate response to this environmental change through an associated pathway (Stock et al. 2000) (Figure 1.1 A). The sensor histidine kinase usually functions as a homodimer, as one monomer catalyses the phosphorylation of the histidine residue on itself or on the other monomer (Stock et al. 2000, West and Stock 2001).

It was then found that TCSs also present in eukaryotic organisms, such as yeast, algae, fungi, and higher plants (Lohrmann and Harter 2002, Oka et al. 2002). In addition, a more complex type of TCS, the multistep system, was found as being mostly in yeast and plants (Wurgler-Murphy and Saito 1997, Stock et al. 2000, Thomason and Kay 2000). In multistep TCS, there is a hybrid histidine kinase carrying an additional

receiver domain at its C-terminus. When the sensor histidine kinase perceives a signal, and being phosphorylated, it firstly transfers the phosphoryl group to a conserved aspartate residue on its own receiver domain, then to a Histidine Phosphotransfer (HPt) protein, and finally to the receiver domain of the response regulator, bringing about the appropriate response to the signal (Figure 1.1 B).

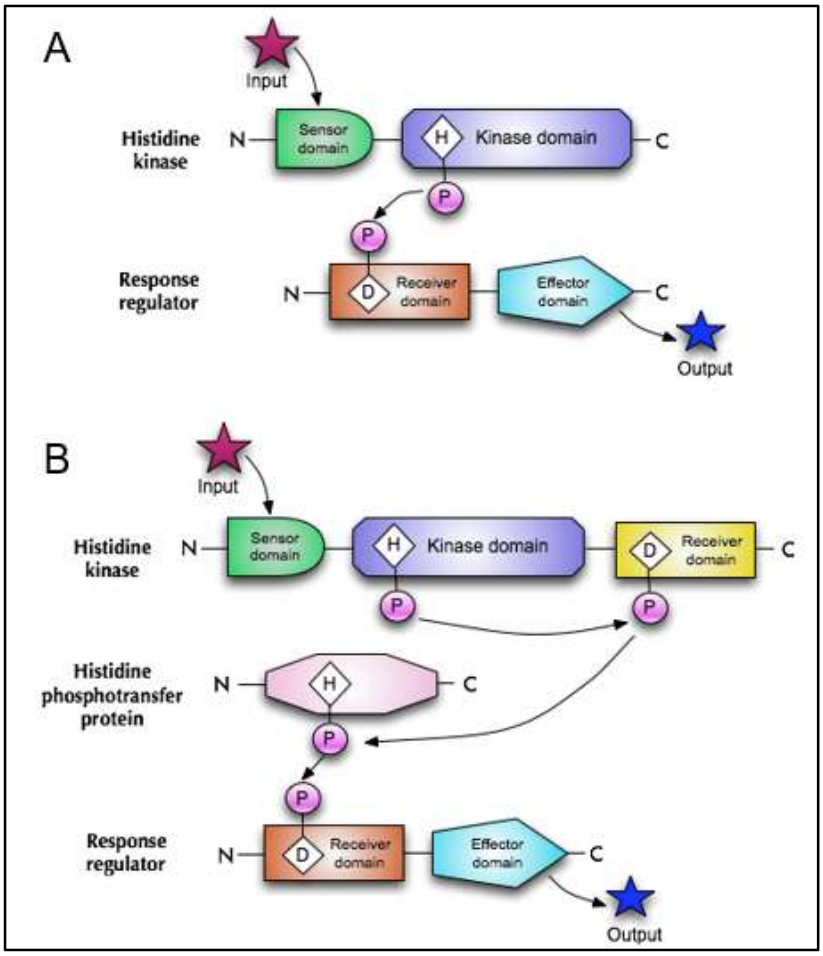


Figure 1.1 General models of single-step TCSs and multistep TCSs. A) The single-step TCS consists of a sensor histidine kinase and a response regulator, the autophosphorylation on the histidine kinase is triggered by the input signal, and the phosphoryl group is subsequently transferred to the response regulator. B) In multistep TCS, the histidine kinase usually carries an additional receiver domain at its C-terminus, and there is another component, a HPt protein. The phosphoryl group transfers from the kinase domain to the receiver domain of the histidine kinase, and then through the HPt protein to the response regulator.

An example of a well-studied TCS in *Escherichia coli* is the EnvZ/OmpR two component system responding to osmotic signals. In the environment with high osmolality, such as 0.5 M NaCl or 20% (w/v) sucrose, EnvZ becomes autophosphorylated and transfers the phosphoryl group to its cognate response regulator OmpR. Then the phosphorylated OmpR binds to the promoter regions of two outer membrane porin genes, *ompF* and *ompC*, differentially modulating their transcription (Comeau et al. 1985, Mattison and Kenney 2002, Yoshida et al. 2002). OmpF and OmpC can form channels in the outer membrane, allowing the passive diffusion of small hydrophilic molecules to adjust the osmotic pressure (Nikaido and Vaara 1985). When the osmotic stress fades away, EnvZ can also dephosphorylate the phosphorylated OmpR, lowering the porin genes expression accordingly (Aiba et al. 1989).

1.1.1 Histidine kinase

Histidine kinases belong to the protein kinase family. A kinase is a form of enzyme that catalyses a phosphorylation reaction, which is the transfer of phosphoryl groups to a substrate protein. The phosphoryl group usually comes from a Nucleoside Triphosphate (NTP), mostly from an Adenosine Triphosphate (ATP). There are 3 major types of protein kinases, the serine/threonine kinases, tyrosine kinases and histidine kinases, respectively phosphorylating on the serine/threonine, tyrosine, and histidine residues of the substrate proteins. Here, we focus on histidine kinases.

1.1.1.1 Membrane-bound histidine kinase

A typical sensor histidine kinase contains two signalling components, a variable sensor domain (input domain) that is unique to each histidine kinase, and an invariable kinase domain (output domain) with highly conserved structure and function (Figure 1.2 A). Most sensor histidine kinases are membrane proteins, so that they can perceive signals outside the cell by the sensor domain and evoke responses inside the cell by the kinase domain, acting as a cellular receptor. Figure 1.2 B is a model of a typical Transmembrane (TM) histidine kinase, exemplified by the well-studied *E. coli* osmosensor EnvZ. There are two TM α-helices (termed TM1 and TM2) at the N-terminus of one monomer, with an extracytosolic sensor domain lying between them. The second TM α-helix is followed by a linking segment HAMP domain, which was given this name because of its presence in Histidine kinases, Adenylyl cyclases, Methyl-accepting chemotaxis proteins and Phosphatase (Aravind and Ponting 1999), then followed by a Dimerisation and Histidine phosphorylation (DHp) domain and a Catalytic and ATP-binding (CA) domain at the C-terminus. As is shown in Figure 1.2 B, the HAMP domain forms two helices, named H(α1) and H(α2), and the DHp domain also forms two helices, named D(α1) and D(α2). They are linked as continuous helices, forming an elongated helical bundle. Two helical bundles, from two histidine kinase monomers, are involved in homodimerisation. The CA domain is attached to D(α2), floating by the side of the helical bundles. DHp and CA domains contain several highly-conserved sequences and structures, and a conserved histidine residue which is responsible for accepting the phosphoryl group.

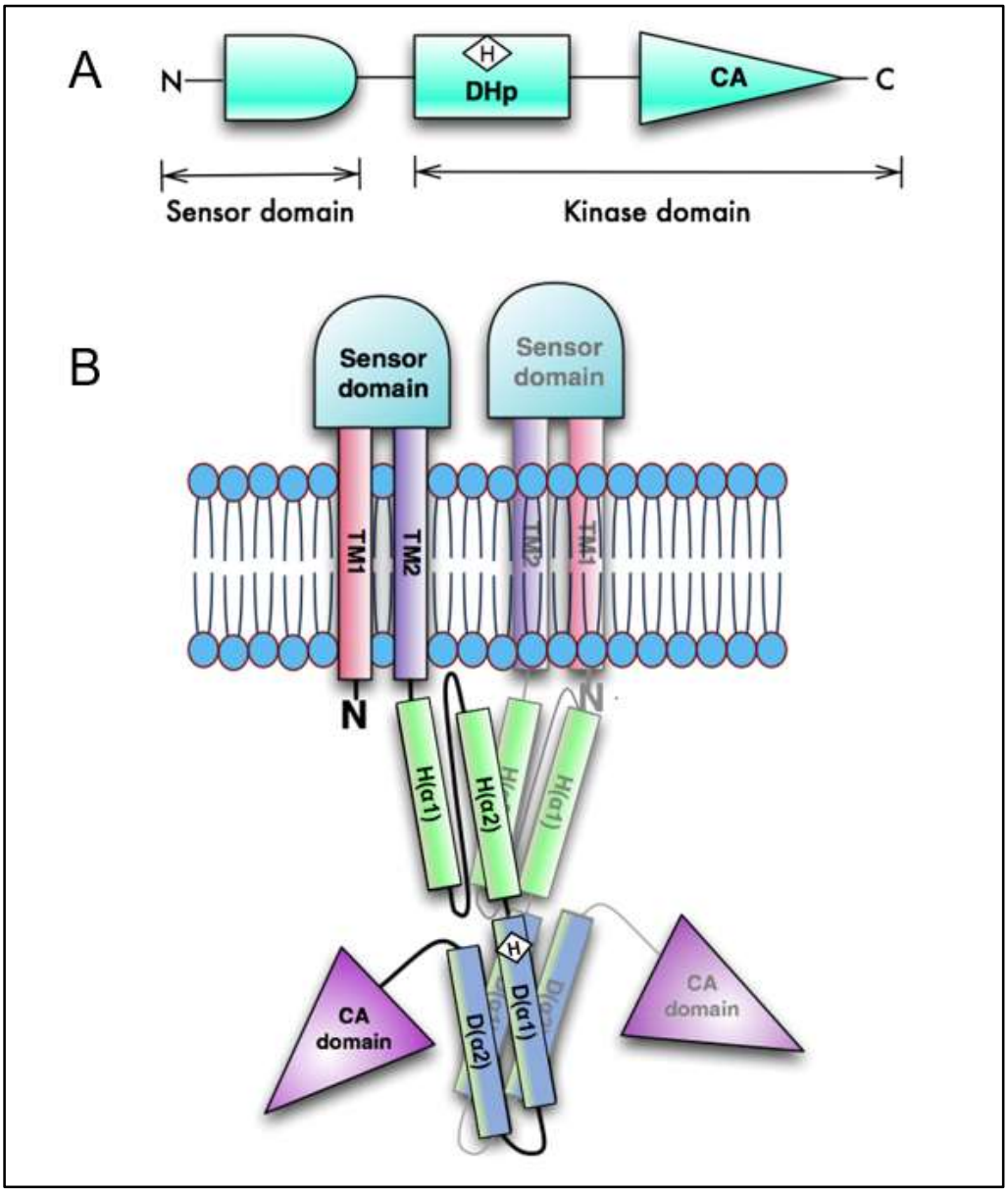


Figure 1.2 Schematic model of sensor histidine kinase structure. A) A schematic structure of a single prototypical sensor histidine kinase. B) Typical transmembrane dimeric histidine kinases structure, consisting of two monomer molecules, exemplified by the *E. coli* osmosensor EnvZ. TM1 and TM2, transmembrane α-helices; H(α1) and H(α2), HAMP domain α-helices; D(α1) and D(α2), DHp domain α-helices. The conserved histidine residue is indicated in rhombus. Panel B was drawn based on Figure 1 from (Wang et al. 2012).

1.1.1.2 Soluble histidine kinase

Membrane histidine kinases comprise a transmembrane sensor domain and a cytoplasmic regulator domain, so that the organisms are able to respond to extracellular signals in the environment. In the meanwhile, there are certain stimuli, such as light and diffusible molecules, which can be sensed by soluble histidine kinases existing in the cytoplasm.

One example is a noncanonical histidine kinase ExsG in *Rhizobium* sp. NT-26 (Wojnowska et al. 2013). ExsG is encoded along with the Bacteriophytochrome Photoreceptor 1 (BphP1), which is a red and far-red light sensor in bacteria, and its counterparts in plants and cyanobacteria control light dependent processes (Davis et al. 1999, Smith 2000, Bhoo et al. 2001, Kumar et al. 2012). Figure 1.3 presents the domain architecture of ExsG, showing that instead of a sensor domain, there is a receiver domain (REC) (a subdomain in response regulators) at the N-terminus, and an HWE_HK domain (defined by the presence of a conserved histidine residue and a WXE motif), which corresponds to the DHp domain in canonical histidine kinases, and a CA domain. ExsG was found to be soluble and form hexamer in solution; this will be discussed in Chapter 3 (Wojnowska et al. 2013). In addition, it has been reported that in *Synechocystis* sp. PCC 6803, there are several histidine kinases identified to be soluble, such as Hik34 (a unique histidine kinase without CA domain), Hik41 (a hybrid histidine kinase, containing a receiver domain), and Hik2, together with their cognate response regulators, functioning as transducers of hyperosmotic stress (Paithoonrangsarid et al. 2004).

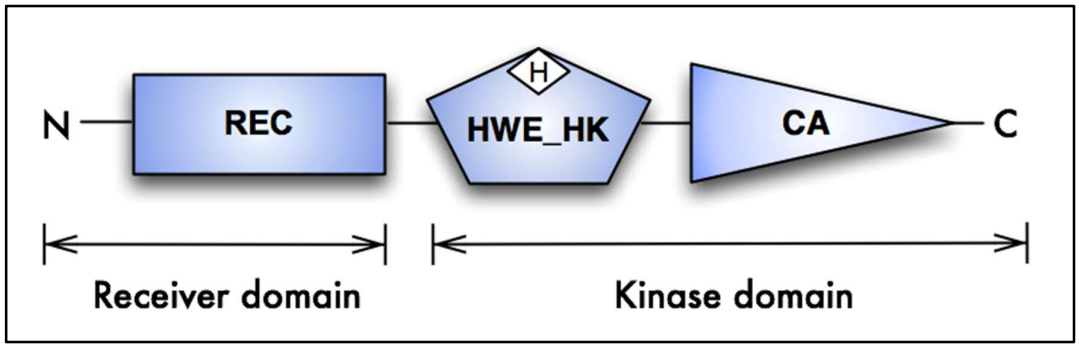


Figure 1.3 Domain architecture of a soluble histidine kinase *Rhizobium* ExsG. Compared to the canonical histidine kinases, ExsG has no sensor domain, but a receiver domain. REC refers to the receiver domain; HWE_HK refers to a HWE domain that defines the HWE histidine kinase family by several conserved residues, a histidine, and a tryptophan-X-glutamic acid sequence; CA refers to catalytic and ATP-binding domain.

1.1.1.3 Sensor domain

Sensor domains can be found located (i) in the membrane (termed 'membrane-embedded sensor'), (ii) inside the cytosol (termed 'intracytosolic sensor') or (iii) outside the cytosol (termed 'extracytosolic sensor'), depending upon the specific signals they sense; there is only a small amount of primary sequence identity shared by these types of sensor domains.

(i) The membrane-embedded sensor domains often have numerous transmembrane segments, containing a sensory region inside their tertiary structure, and lack extracellular domains. Currently, little structural information has been obtained, but so far it has been demonstrated that the membrane-embedded sensor domains function in the detection of ethylene (Voet-van-Vormizeele and Groth 2008), thermal-response (Albanesi et al. 2009, Martin et al. 2009) and redox-sensing (Bogel et al. 2009).

(ii) Many intracytosolic sensor domains have related structural folds, and can be classified as having a PAS domain (named from its presence in Period circadian proteins, Aryl hydrocarbon receptor nuclear translocator proteins and Single-minded proteins) (Taylor and Zhulin 1999, Hefti et al. 2004) or a GAF domain (named from its presence in cGMP-specific phosphodiesterases, Adenylyl cyclases and the bacterial transcriptional regulator Formate hydrogenlyase A (FhlA)) (Aravind and Ponting 1997, Martinez et al. 2002). The amino acid sequences of GAF and PAS domains are unrelated, but through the determination of a number of X-ray structures from different organisms, similarities regarding their structural topologies have been revealed. It has been shown that GAF and PAS domains can sense light (Wagner et al. 2005, Avila-Perez et al. 2006), O_2 (Gong et al. 1998, Roberts et al. 2004), small ligands (Bruder et al. 2005, Cho et al. 2006) and redox potential (Quinn et al. 2011), and regulate signalling pathways in diverse organisms (discussed below). However, the functions of most GAF and PAS domains remain unresolved.

Since Hik2 has a GAF domain as its sensor domain, as predicted here by the online server SMART (Simple Modular Architecture Research Tool) (Schultz et al. 1998, Letunic et al. 2015), placing it within the intracytosolic family, it is worth further discussing about how the GAF domain acquires the sensor function, and then relating this to the PAS superfamily.

GAF domain

As seen from Figure 1.4, the GAF domain consists of a central six-stranded antiparallel β-sheet core, and an outer layer of a mixture of short α-helices and loops. There are

numerous GAF domain structures being determined, and are often found in tandem form.

For example, in cyanobacterial adenylyl cyclase CyaB2, there are two tandem GAF domains, crystallised as a homodimer in an antiparallel manner (Martinez et al. 2005) (Figure 1.4 A). Each monomer consists of two GAF domains, GAF A, and GAF B, connected by a long α-helix, which is mainly responsible for the dimerisation. GAF A and GAF B share a similar cAMP (cyclic adenosine monophosphate) binding site and a similar structure, except for the loop connecting β2 and β3 strands. Both GAF domains can bind cAMP, allowing CyaB2 perform a function of increasing the cyclase activity and the cAMP synthesis (Bruder et al. 2005).

*Dr*BphP is a phytochrome in the proteobacterium *Deinococcus radiodurans*, containing both GAF and PAS domains as its N-terminal chromophore-binding domain. A crystal structure has revealed that the GAF domain binds a bilin chromophore biliverdin by the cysteine residue 24, in a deep solvent-shielded pocket (Wagner et al. 2005, Wagner et al. 2007) (Figure 1.4 B). The biliverdin-bound *Dr*BphP can be converted by light between Pr and Pfr state, which absorbs light at maxima 698 and 750 nm respectively, functioning as a light-regulated kinase to optimise the shade detection (Bhoo et al. 2001, Wagner et al. 2007).

DosT is a sensor histidine kinase identified in *Mycobacterium tuberculosis*, together with another histidine kinase DevS and the response regulator DosR, detecting reduced O_2 tension and the concentration of NO and CO exposures via the heme-binding GAF domains (Roberts et al. 2004) (Figure 1.4 C). The CO- or NO-bound DosT and deoxy

DosT has a significant increasing activity of autokinase, compared to the oxy DosT, thus, DosT may function as a hypoxia sensor (Kumar et al. 2007). DosT contains two tandem GAF domains. Although the GAF domain in DosT contains a five-stranded antiparallel β-sheet instead of six-stranded, the overall structure and the heme-binding mode is identical to the GAF domain (Podust et al. 2008).

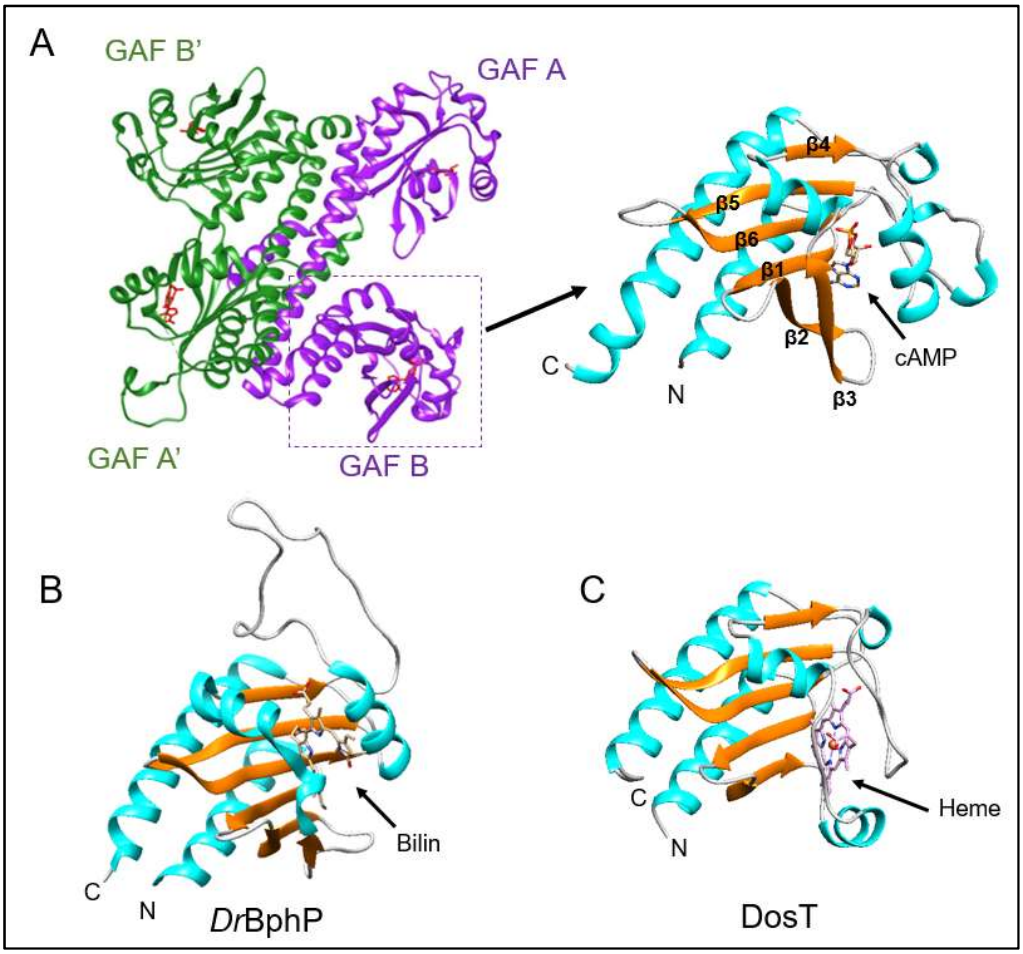


Figure 1.4 Structures of the GAF domain in sensor families most related to Hik2. A) Homodimer of the tandem GAF domain and the amplified GAF B in CyaB2 (PDB code: 1YKD, *Anabaena* PCC7120) (Martinez et al. 2005). B) The GAF domain of *Dr*BphP (PDB code 2O9C, *Deinococcus radiodurans*) (Wagner et al. 2007). C) The GAF domain of DosT (PDB code: 2VZW, *Mycobacterium tuberculosis*) (Podust et al. 2008). The models are aligned by their central β-sheet cores. α-helices are rendered in cyan, and β-strands in orange. The molecules bound to the GAF domains are indicated by arrows. This figure was drawn using UCSF Chimera (Pettersen et al. 2004).

PAS domain

Similar to the GAF domain, the PAS domain consists of a central five-stranded antiparallel β-sheet core, flanked by α-helices. It is composed of around 100 to 120 amino acids and is commonly found in sensor histidine kinases, methyl-accepting chemotaxis proteins and cGMP synthases/hydrolases.

For example, FixL is a typical histidine kinase containing a PAS domain and a kinase domain. FixL binds *b*-type heme cofactor via its PAS domain, functioning as a phosphodiesterase regulated by oxygen. Crystal structure information reveals that the heme is coordinated by a histidine residue at sequence position 200 as the ligation site in the PAS domain (Gong et al. 1998) (Figure 1.5).

Another example is the LOV (light-, oxygen-, or voltage) domain, a member of the PAS domain superfamily, which can bind the flavin cofactor Flavin Mononucleotide (FMN) via a cysteine residue, and respond to the absorption of blue light directly. LOV domain was first identified in NPH1 in *Arabidopsis thaliana* as a photosensory domain (Huala et al. 1997), and then in fungi such as in VVD in *Neurospora crassa* (Heintzen et al. 2001), and in bacteria such as YtvA in *Bacillus subtilis* (Avila-Perez et al. 2006) (Figure 1.5).

There are also a few PAS domains being in the extracytosolic sensor group, such as the one in PhoQ, which belongs to the PDC fold (named from its presence in the first determined structures of the type of sensor domains in proteins PhoQ, DcuS and CitA). PhoQ is a conserved histidine kinase and can be found in a variety of gram-negative pathogens, regulating virulence, and responding to low pH and low divalent metal ion

signals together with its cognate response regulator PhoP (Prost and Miller 2008). Under different environmental conditions, a variety of divalent metal ligands can be bound by the PAS domain, including Mg^{2+} (Garcia Vescovi et al. 1996), Ca^{2+} (Cho et al. 2006) and Ni^{2+} (Cheung et al. 2008). In Figure 1.5, the structural information of the PAS domain in *Salmonella typhimurium* PhoQ shows that the Ca^{2+} ion binds to a surface made of a cluster of acidic residues which is close to the inner membrane (Cho et al. 2006).

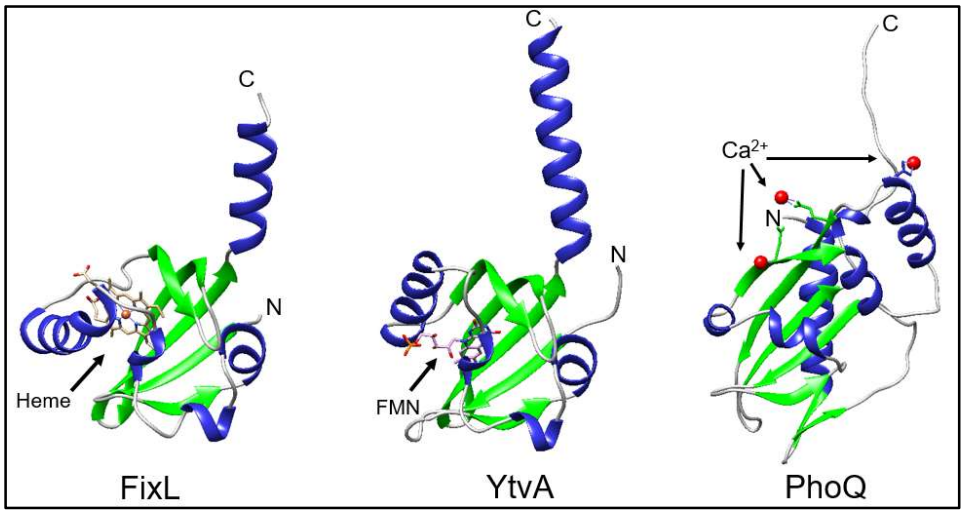


Figure 1.5 Structures of the PAS domain. The GAF domain in FixL (PDB code: 1DRM, *Bradyrhizobium japonicum*) (Gong et al. 1998); the YtvA (PDB code: 2PR5, *Bacillus subtilis*) (Moglich and Moffat 2007); the PhoQ (PDB code: 1YAX, *Salmonella typhimurium*) (Cho et al. 2006). The models are aligned centrally by their β-sheet cores. α-helices are rendered in blue, and β-strands in green. The molecules bound to the PAS domains are indicated by arrows. This figure was drawn using UCSF Chimera (Pettersen et al. 2004).

(iii) Unlike the intracytosolic sensor domain, the sequences of extracytosolic sensor domains are highly diverse. However, according to the known structure information, most of the extracytosolic sensor domains belong to three structural types: all α-helix, β-sandwich and mixed α/β fold.

All α-helix type sensor domains have been reported in NarX in *Escherichia coli* (Cheung and Hendrickson 2009) and TorS in *Vibrio parahaemolyticus* (Moore and Hendrickson 2009). NarX has a few relatively short α-helices, while TorS has six helices, forming two four-helix bundles (Figure 1.6 A). Two monomers of the NarX sensor domain form a dimer of four-helix bundles, with a single nitrate ion partially buried between the two α-helices in the dimer interface (Cheung and Hendrickson 2009). TorS associates with its co-receptor TorT, forming a sensor complex. The osmoregulator Trimethylamine-N-oxide (TMAO) binds to TorT and sends a signal through the long TorS α-helices, stimulates the phosphorylation of TorS, functioning in a TMAO-based respiration (Moore and Hendrickson 2012).

For the β-sandwich type, the RetS in *Pseudomonas aeruginosa* is the only crystal structure solved (Jing et al. 2010) (Figure 1.6 B). There are two opposing antiparallel β-sheets in the centre, surrounded by loops and short α-helices. Although the signals that regulate RetS are not known, it is believed that its sensor domain perceives an environmental signal, causing the RetS homodimer to dissociate to form a heterodimeric complex with another signalling kinase GacS, thereby regulating the transcription of genes associated with cytotoxicity and acute infections (Jing et al. 2010).

The third type of extracytosolic sensor domain is mixed α/β fold, with a typical example is the PDC fold, which shares a similar structure with the intracytosolic sensor, the PAS domain. The PDC fold sensor domain responds to extracellular divalent cations in PhoQ (Garcia Vescovi et al. 1996), certain C4-dicarboxylates in DcuS (Golby et al. 1999), and citrate in CitA (Kaspar et al. 1999). As shown in Figure 1.6 C, the PDC fold often

begins with a long N-terminal helix, followed by a five-stranded antiparallel β-sheet in the centre, flanked by a few small α-helices, and ends with a very short helix at its C-terminus.

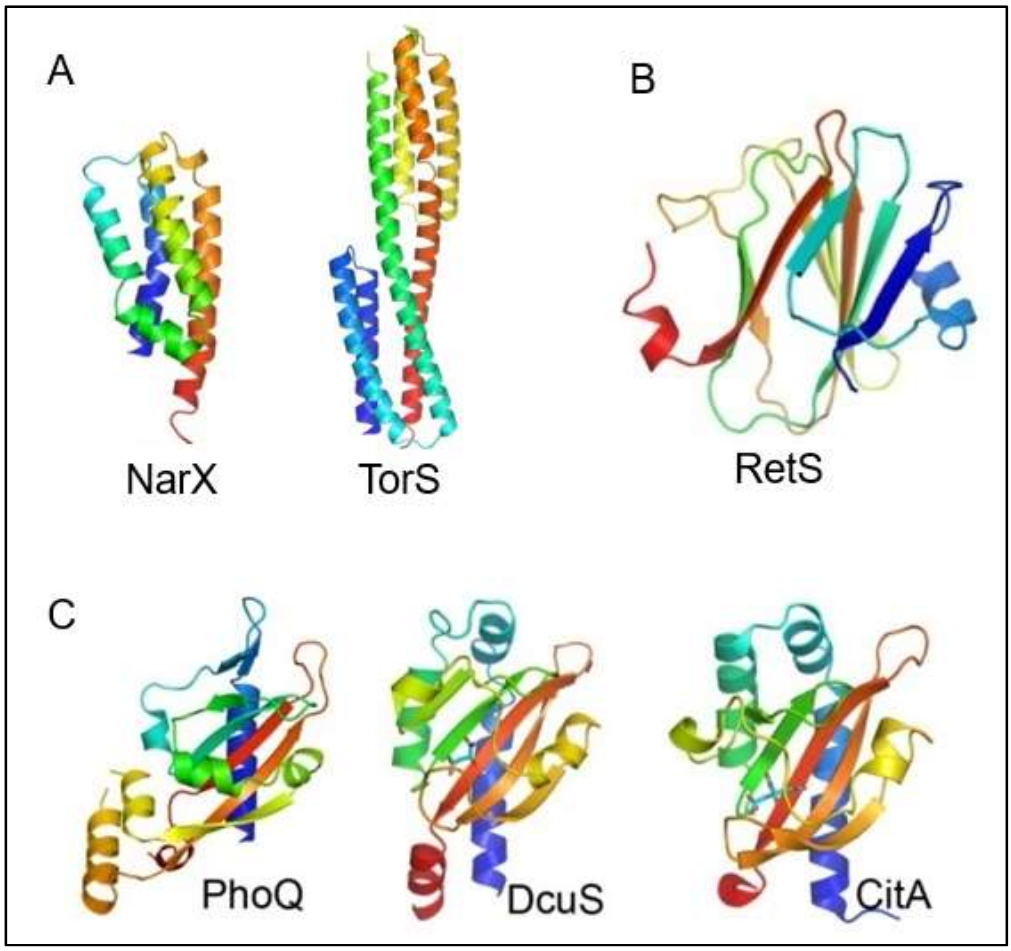


Figure 1.6 Structures of three types of extracytosolic sensor domains. A) All α-helix type in NarX (PDB code: 3EZH, *Escherichia coli*) (Cheung and Hendrickson 2009) and TorS (PDB code: 3I9Y, *Vibrio parahaemolyticus*) (Moore and Hendrickson 2009). B) β-sandwich type in RetS (PDB code: 3JYB, *Pseudomonas aeruginosa*) (Jing et al. 2010). C) Mixed α/β type in PhoQ (PDB code: 3BQ8, *Escherichia coli*) (Cheung et al. 2008), DcuS (PDB code: 3BY8, *Escherichia coli*) (Cheung and Hendrickson 2008) and CitA (PDB code: 2J80, *Klebsiella pneumoniae*) (Sevvana et al. 2008). The structures are each coloured in a rainbow spectrum from blue to red from the N- to C-terminus, respectively. Figure adapted from (Wang 2012).

The PDC fold was originally distinguished from the PAS domain according to the difference of the residues in the N-terminal helix and between the second and third β-strands (Cheung et al. 2008). Besides, the PDC fold has a 3α-2β-α-3β or 2α-3β-α-3β topology, while the PAS domain has an α-2β-4α-3β topology (Moglich et al. 2009, Chang et al. 2010, Scheu et al. 2010) (Figure 1.7). However, there exists literature which defines the PDC fold as a subclass of the extracytoplasmic PAS sensor domain, due to the similarity of the overall structure observed (Shah et al. 2013).

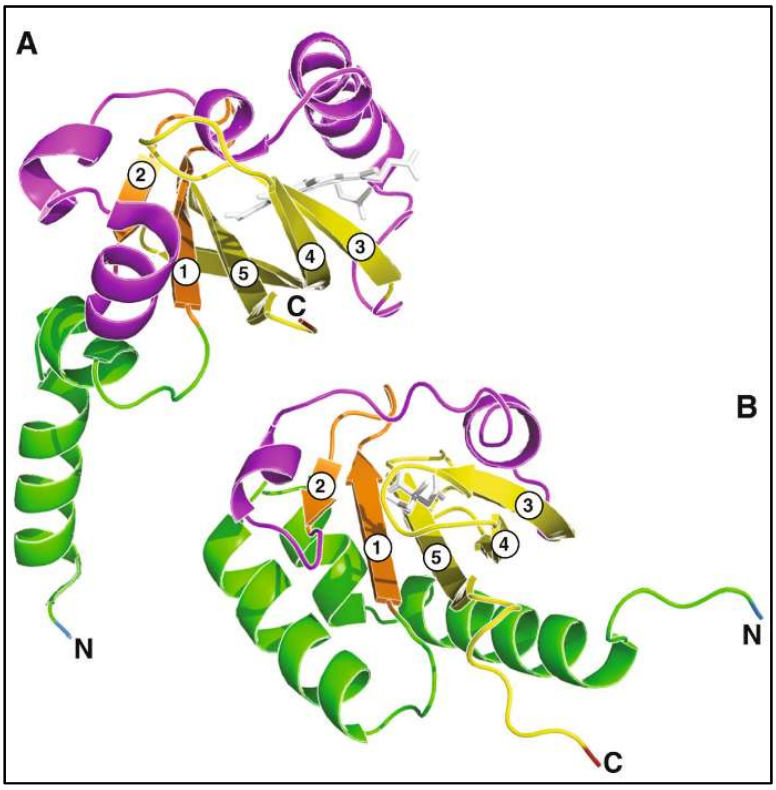


Figure 1.7 The comparison of 3D-structures of a PAS domain and a PDC fold. A) The PAS domain in FixL with its ligand heme in *Rhizobium meliloti* (PDB code: 1D06). B) The PDC fold in CitA with its ligand citrate in *Klebsiella pneumoniae* (PDB code: 1P0Z). For both structures, the core β-strands are labelled from 1 to 5. Each region is coloured as follows: the amino end with blue, the leading α-helix region with green, the first two β-strands with orange, the inter-domain α-helix region with magenta, the last three β-strands with yellow, and the carboxyl end with red. Figure adapted from (Shah et al. 2013).

Both the PAS and PDC domains are of the globular mixed α/β fold, as well as the sensor domain of Hik2, a GAF domain. The topology of the five-stranded β-sheet core in PAS domain/PDC fold corresponds to the six-stranded one in GAF domain, and there are also loops and α-helices flanked by the side. Thus, describing the structures of all the above sensor domains may give us clues when speculating about the function of the GAF sensor domain in Hik2.

1.1.1.4 Kinase domain

The kinase domain is the catalytic core of a histidine kinase, consisting of two conserved subdomains, a DHp domain (referred to HisKA in Pfam database (Finn et al. 2016)) and a CA domain (referred to HATPase_c in Pfam database), connected via a short linker. As shown in Figure 1.9 C, the two CA domains are flanking the four-helix bundle of dimeric DHp domains on both sides, assembling to a kinase domain dimer, which carries the catalytic properties in histidine kinases. Both of the two domains have signature motifs (boxes) termed on the basis of specific sequences: the H box in the DHp domain, which is the site for autophosphorylation; the N, G1, F, G2 and G3 boxes in the CA domain, for binding ATP and donating the γ-phosphate group of the ATP to the conserved histidine residue on the DHp domain (Figure 1.8).

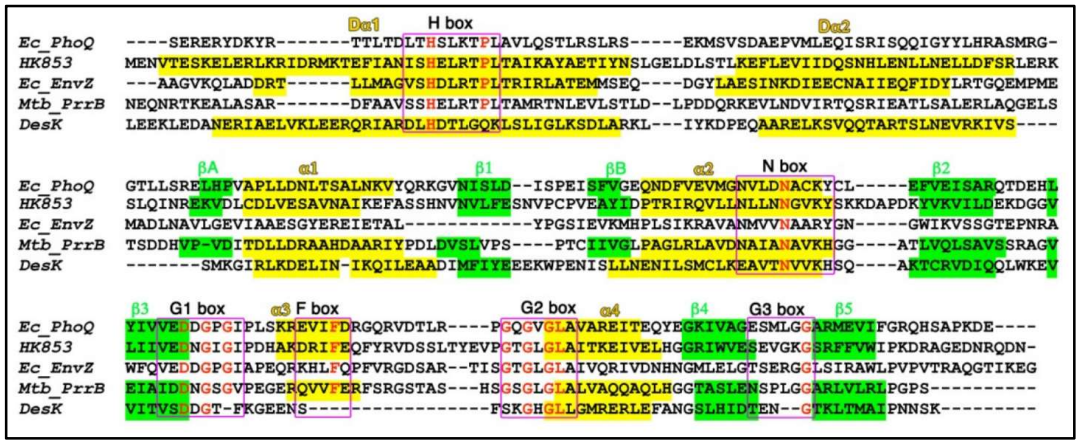


Figure 1.8 Sequence alignment of the kinase domains within histidine kinases. Conserved motifs H, N, G1, F, G2 and G3 boxes are in red boxes. α-helices are highlighted in yellow, and β-strands in green. Figure adapted from (Wang 2012).

The DHp domain consists of two α-helices (named Dα1 and Dα2) forming an antiparallel coiled-coil, and a pair of DHp domains usually associate to form a four-helix bundle (Figure 1.9 A), which is responsible for the dimerisation of histidine kinases. The conserved histidine residue is located on the first α-helix (Dα1), which is more mobile than Dα2; and this phosphorylation site is solvent exposed, becoming convenient for the acceptance of a phosphoryl group, and transfer of signals.

The structure of a CA domain is demonstrated in Figure 1.9 B, using the C-terminal PhoQ catalytic domain that was crystallised with the non-hydrolysable ATP analog AMPPN as an example (Marina et al. 2001). The CA domain is generally globular, existing as a monomer. It is an α/β sandwich fold with two layers, one containing three α-helices and the other one having five β-strands. The signature motifs N, G1, F, G2 and G3 boxes are involved in binding ATP. There is a flexible loop region named ATP lid between F box and G2 box, covering the ATP-binding site. When binding an ATP analog, the ATP lid covers the nucleotide, having a relatively ordered structure (Bilwes

et al. 2001, Marina et al. 2001), while in the absence of ATP, the ATP lid is highly flexible in the crystal structures (Bilwes et al. 1999, Song et al. 2004). The conformational change of this ATP lid allows the CA domain to associate with the DHp domain at different points, thus acting as three different enzymes, autokinase, phosphotransferase, or as a phosphatase.

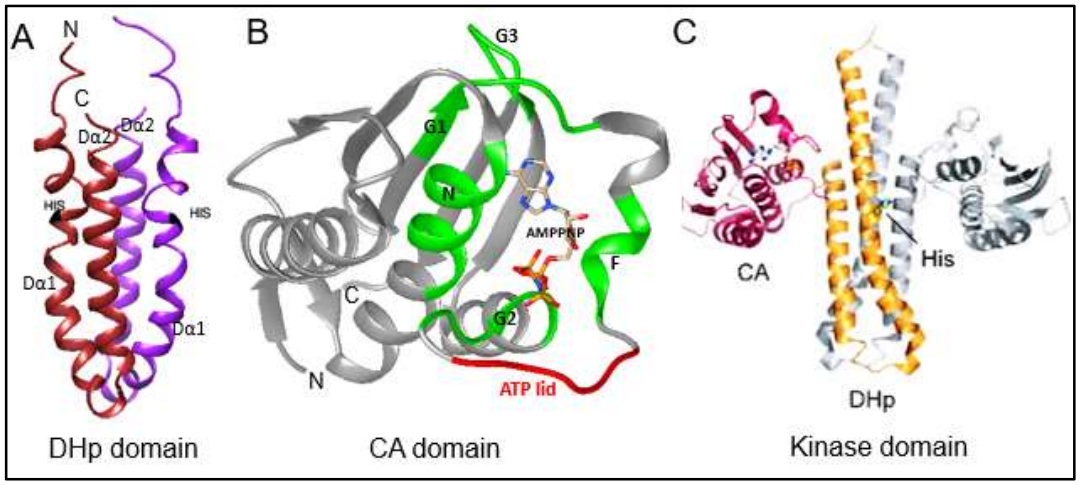


Figure 1.9 Structures of subdomains for the kinase core of histidine kinases. A) The DHp domain of EnvZ (PDB code: 1JOY, *Escherichia coli*) (Tomomori et al. 1999). The conversed phosphoacceptor histidine residue is labelled in black. B) The CA domain of PhoQ (PDB code: 1ID0, *Escherichia coli*) (Marina et al. 2001). Homology boxes N, G1, F, G2 and G3 are coloured in green, and the ATP lid in red. The ATP analog AMPPNP is shown in stick. C) The dimeric kinase domain of HK853 (PDB code 2C2A, *Thermotoga maritima*) (Marina et al. 2005, Gao and Stock 2009). One monomer is coloured in orange and pink, and the other in grey. This figure was prepared using UCSF Chimera (Pettersen et al. 2004).

1.1.1.5 Signal transduction mechanism

In a histidine kinase, the sensor domain senses the environmental signal and the kinase domain transfers this signal to response regulator. A question arises as to how the signal transfers from the sensor domain to the kinase domain, allowing the CA domain to perform different catalytic functions. Since the prototypical histidine kinase is a transmembrane protein, a well-characterised model is shown in Figure 1.10, describing a ligand-binding induced conformational change.

The transmembrane histidine kinase has two antiparallel transmembrane helices termed TM1 and TM2 which connect the extracellular sensor domain to the cytoplasmic part. When the sensor domain receives a signal, subtle piston-like or rotation movements, or changes combining both, are induced on the transmembrane helices. This causes a structural movement of the kinase domain dimer, either directly or relayed by an HAMP domain that links transmembrane helices and DHp domains. This movement results in a rotation or tilting of the cytoplasmic helix, causing conformational changes of the two helices in DHp dimer packing and in the interface between DHp and CA domains, and the subsequently coordinated displacement of the CA domains, such as its distance and relative position to the DHp domain. In this way, the disposition of the CA domain affects the activity of the ATP-binding site, by either aligning the ATP molecule with the phosphoacceptor site on DHp domain to perform the autophosphorylation reaction, or presenting a docking site for the response regulator to transfer the phosphoryl group to perform the phosphotransfer and phosphatase reactions, as discussed in section 1.1.1.4 (Casino et al. 2010).

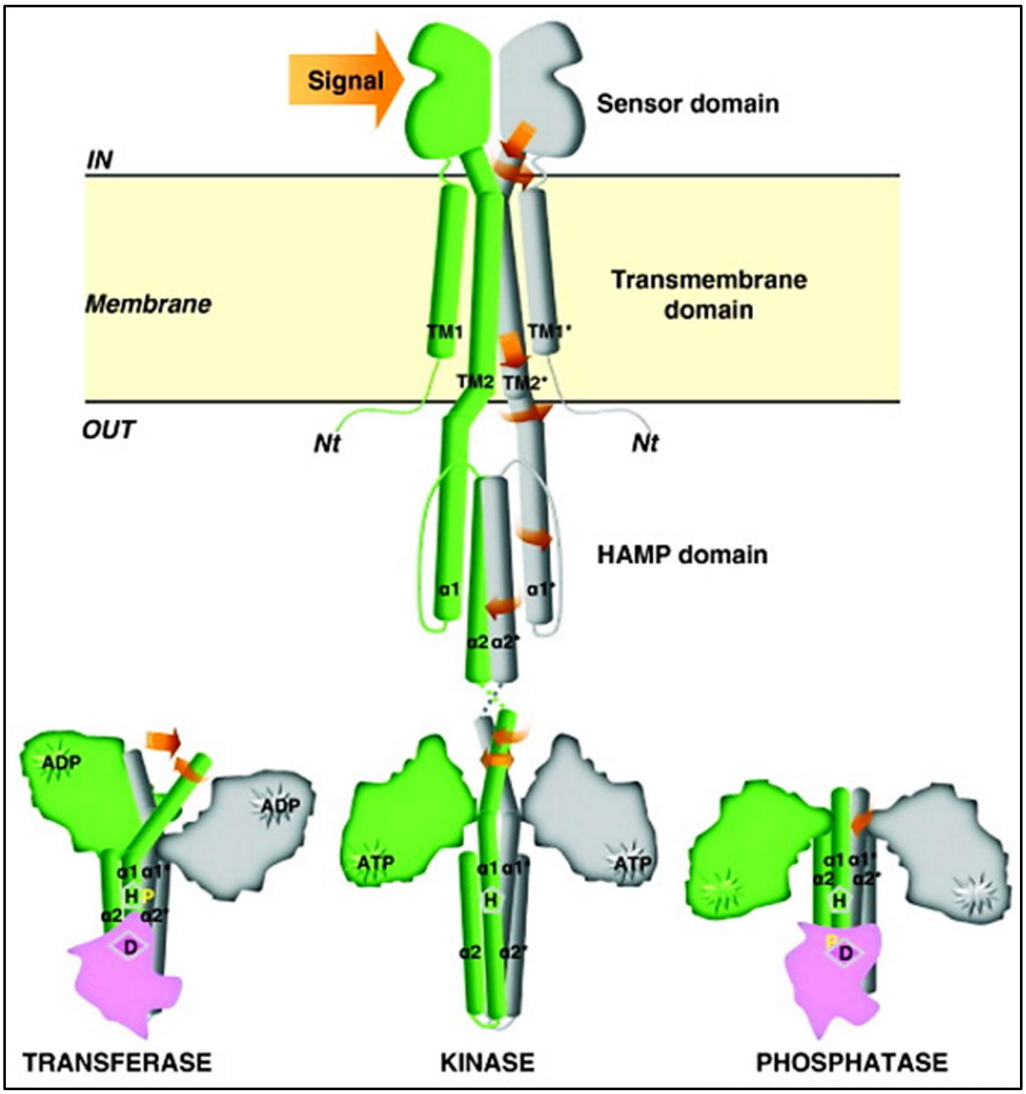


Figure 1.10 The model of signal transduction in histidine kinase. When the sensor domain receives signal, subtle piston-like or rotation changes are induced in the TM2 helix. The Dα1 in DHp domain that coupled TM2 moves concomitantly, either directly or through an HAMP domain that links TM2 and Dα1 via a rotation of the connecting helices, causing conformational changes of the two helices in DHp dimer packing, making CA domain to move close to or away from the DHp domain. Thus, the three functional states of a histidine kinase are determined by the different relative disposition of CA and DHp domains, either aligning the ATP molecule with the phosphoacceptor site on DHp domain, or presenting a docking site for the response regulator, which affects the catalytic activity of the CA domain. Figure from (Casino et al. 2010).

For the cytoplasmic histidine kinase, such as Hik2, the signalling mechanism is not clear yet, but the same output and modulation can be obtained from differing conformational changes. Analysis of the transition from apo-protein to ligand-bound protein state represents a transition between symmetric and asymmetric states is required, which has indeed been described in structures of sensor domains in Tsr, TorS, LuxQ, and DctB proteins (Neiditch et al. 2006, Zhou et al. 2008, Moore and Hendrickson 2012, Winkler et al. 2012).

1.1.2 Response regulator

The response regulator is the second component within the classical two-component system. The majority of response regulators act as a transcription factor, which accept phosphoryl groups from their cognate histidine kinases and elicit genetic responses (Gao and Stock 2009); a few can participate in certain enzymatic activities, such as the chemotaxis methylesterase CheB (Anand et al. 1998) and the cAMP-phosphodiesterase RegA (Shaulsky et al. 1998). It has been reported by recent research that Hik2 has two identified cognate response regulators Rre1 (Response regulator 1) and RppA (Regulator of photosynthesis and photopigment-related gene expression A), which interact with Hik2 directly and function with Hik2 in an important signalling pathway (Ibrahim et al. 2016).

There are two domains in a prototypical response regulator, an N-terminal receiver domain that accepts the phosphoryl group form its cognate histidine kinase, and a C-terminal effector domain that generates output response. The structure of the receiver domain is quite conserved, functioning as a switch activated by phosphorylation. However, the effector domains from different response regulators are variable, so that

they are able to provide diverse output functions. Also, some response regulators containing only receiver domain, transfer signals by interacting with other proteins, such as the CheY protein in the chemotaxis system in *E. coli* (Stock et al. 1989), and the sporulation factor Spo0F in *B. subtilis* (Appleby et al. 1996).

1.1.2.1 Receiver domain

The FixJ protein in the symbiotic bacterium *Sinorhizobium meliloti* is a typical response regulator, which activates the transcription of nitrogen-fixation genes *nif* and *fix* when being phosphorylated (David et al. 1988). The receiver domain structure of FixJ is shown as an example in Figure 1.11 A. It is a mixed α/β fold, which is made of a central five-stranded parallel β-sheets and five α-helices flanking on both sides, α1 and α5 on one side, and α2, α3 and α4 on the other side. Most conserved active site residues are located at β1, β3 and β4 strands at the C-terminal end. Two highly conserved Aspartic acid (Asp) residues (the first may sometimes be a Glutamic acid residue(Glu)) at the end of β1 are essential for the divalent metal ion binding, which is important for all phosphoryl group chemistry; and a highly-conserved Asp residue at the end of β3 is the phosphorylation site. β4 ends with a conserved Threonine/Serine (Thr/Ser) residue, which interacts with the phosphoryl group (Gao and Stock 2009). This overall structure is well-conserved among response regulators. During the phosphoryl group transfer reaction, the helix α1 is inserted into the two helices in the DHp domain of a histidine kinase, together with the loop between β5 and α5, allowing the conserved Asp residue to be aligned with the conserved Histidine (His) residue and the subsequent phosphoryl transfer to occur (Casino et al. 2009).

It has been revealed by structural studies that the receiver domain conformational changes induced by phosphorylation usually occur at the surface of α4-β5-α5, together with movements of two conserved 'switch residues', a Thr/Ser at the C-terminus of β4 and a Tyrosine/Phenylalanine (Tyr/Phe) in the middle of β5. The α4-β5-α5 surface also contributes to the dimerisation of response regulators. Figure 1.11 B uses the superimposed active and inactive form of the receiver domain of PhoB as an example, to illustrate the subtle conformational differences (Bachhawat et al. 2005). In the inactive receiver domains (coloured in purple), the side chains of the switch residues are oriented away from the active site, with the Tyr/Phe exposed to the surface of the α4-β5-α5 face. When the phosphoacceptor Asp residue is being phosphorylated, the side chains of the switch residues move closer to the active site, while the hydroxyl of the Thr/Ser forms a hydrogen bond with a phosphate oxygen and the aromatic ring of the Tyr/Phe. The loop between β4 and α4 and the helix α4 shifts and rotates, allowing the conformation of the Tyr/Phe residue being a more energetically favourable state. Moreover, a conserved Lysine (Lys) residue in the β5-α5 loop forms a salt bridge with the phosphoryl group, which helps to stabilise the structure of the active site (Bourret 2010).

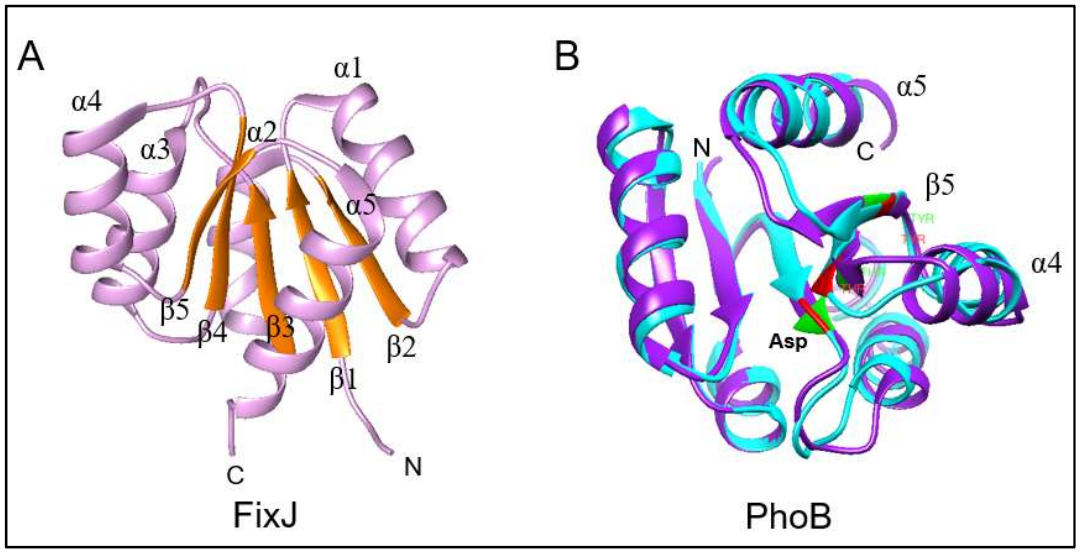


Figure 1.11 Structures of receiver domains in response regulators. A) The structure of the receiver domain in FixJ (PDB code: 1D5W, *Sinorhizobium meliloti*) (Birck et al. 1999). α-helices are rendered in plum, and β-strands in orange. B) The structures of inactive (purple) and active (cyan) receiver domain in PhoB (PDB codes: 1B00 and 1ZES, *Escherichia coli*) (Sola et al. 1999, Bachhawat et al. 2005, Gao and Stock 2009). 'Switch residues' and the phosphoacceptor Asp residue are labelled in red on the active form and green on the inactive form. This figure was drawn using UCSF Chimera (Pettersen et al. 2004).

1.1.2.2 Effector domain

As discussed before, there are single-domain response regulators that contain no effector domain, which take up around 17% of response regulators in bacteria and 50% in archaea. These response regulators usually control bacterial cellular motors, participate in signalling phosphor-relays, and in protein-protein interactions that regulate cell development and division (Jenal and Galperin 2009). The rest of response regulators, with effector domains, have various functions, for example, the binding of

DNA, RNA, proteins, or other ligands, and acting as transporters or enzymes, all of which determine the diversity of their structures. (Galperin 2010).

Most bacterial response regulators contain a DNA-binding domain as their effector domains, making them transcription regulators. The three most common subfamilies are as follows: OmpR/PhoB (30% of all response regulators, containing a winged helix motif, and can be found in all bacterial phyla); NarL/FixJ (17%, containing a helix_turn_helix motif, and can be found in most bacterial phyla except for Chlamydiae, Thermotogae); and LytR/AgrA (3%, mostly containing β-strands, and can be found in bacterial phyla except for Chlamydiae, Chlorobi, Chloroflexi, Cyanobacteria) (Galperin 2010) (Figure 1.12).

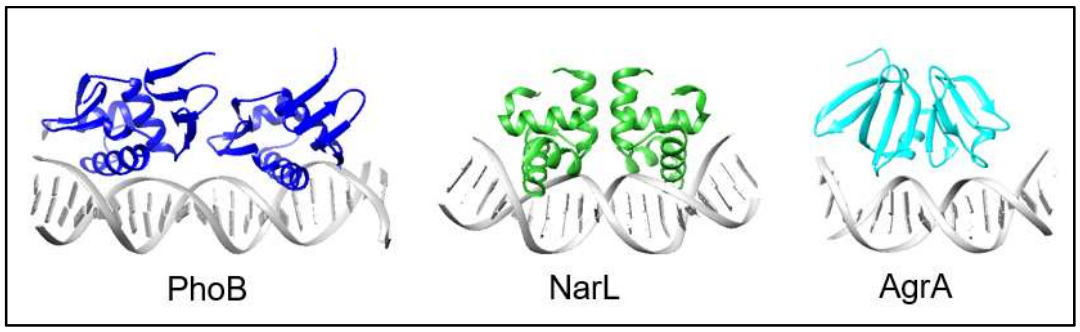


Figure 1.12 Structures of DNA-binding effector domains of response regulators. The structures of the effector domain in PhoB (PDB code: 1GXP, *Escherichia coli*) (Blanco et al. 2002), NarL (PDB code: 1JE8, *Escherichia coli*) (Maris et al. 2002), and AgrA (PDB code: 3BS1, *Staphylococcus aureus*) (Sidote et al. 2008). This figure was drawn using UCSF Chimera (Pettersen et al. 2004).

Recent research revealed Hik2 had two identified cognate response regulator partners Rre1 (Locus ID: slr1783) and RppA (Locus ID: sll0797) (Ibrahim et al. 2016). The domain architecture of Rre1 and RppA are analysed by the SMART online server (Schultz et al. 1998, Letunic et al. 2015) and is shown in Figure 1.13. Rre1 is a NarL

subfamily protein with a helix_turn_helix, Lux Regulon as its effector domain. It has been suggested that Rre1 responds to salt stress and regulate osmotic responsive genes in cyanobacteria (Kaneko et al. 1996, Paithoonrangsarid et al. 2004, Kanesaki et al. 2012), and its homologue Ycf29 in red algal is also known as a NarL-like transcriptional regulator and being the cognate response regulator of CSK (Puthiyaveetil et al. 2010). RppA is an OmpR subfamily protein, containing a transcriptional regulatory domain. It regulates the expression of photosynthesis and photopigment-related genes in cyanobacteria (Li and Sherman 2000).

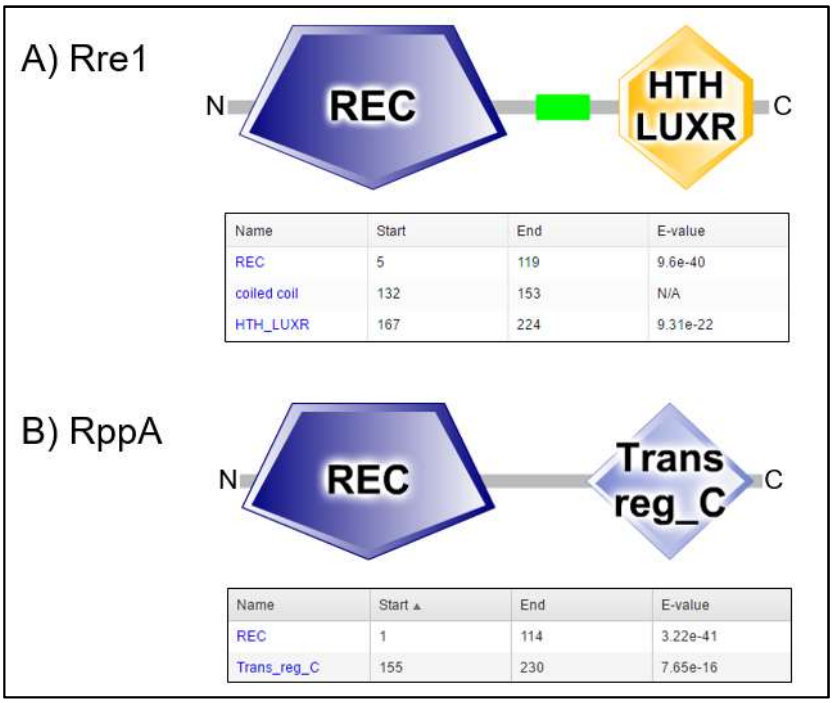


Figure 1.13 Domain prediction of Rre1 and RppA in *Synechocystis* sp. PCC 6803. A) The HTH_LUXR domain refers to the helix_turn_helix, Lux Regulon, which is a DNA-binding, helix-turn-helix domain of about 65 amino acids, present in transcription regulators of the LuxR/FixJ family of response regulators. B) The Trans_reg_C domain refers to the C-terminus of the transcriptional regulatory protein, which is almost always found associated with the response regulator receiver domain, and it may play a role in DNA binding. REC domain, the conserved receiver domain in response regulators. This prediction was carried out using the SMART online server (Schultz et al. 1998, Letunic et al. 2015) for identifying possible signalling domains.

1.2 Histidine Kinase 2

1.2.1 Distribution of Hik2

Histidine Kinase 2 (Hik2) in cyanobacteria has been identified as the homologue protein of *Arabidopsis* Chloroplast Sensor Kinase (CSK), and the genes for Hik2 homologues can be found in all cyanobacterial species for which genomic sequences have been determined (Peter et al. 2015). There are two types of Hik2, Class I and Class II, depending on the domain architecture of the protein as predicted with SMART online server (Schultz et al. 1998, Letunic et al. 2015). Class I is the full-length Hik2, consisting of a sensor domain and a kinase domain, while Class II Hik2, the truncated form, only consists of a kinase domain, but no sensor domain. The distribution of the two types of Hik2 in organisms among cyanobacteria, alga, moss, and plant is shown in Table 1.1 (Ibrahim et al. 2017), and it can be seen that the Class II Hik2 only exists in three cyanobacterial species. Moreover, amongst the Class I Hik2s, most of the sensor domains are predicted to be the GAF domain or a possible GAF domain using the SMART online server. The wide distribution of Hik2 indicates the important role that it plays.

Species	Domain architecture of Hik2		
	Full-length	Truncated	Sensor domain prediction
Synechocystis sp. PCC 6803	✓		GAF
Cyanothece sp. PCC 7822	✓		GAF
Cyanothece sp. PCC 8802	✓		GAF
Cyanothece sp. PCC 8801	✓		GAF
Cyanothece sp. PCC 7424	✓		GAF
Cyanothece sp. ATCC 51142	✓		GAF
Cyanothece sp. PCC 7425	✓		GAF
Crocosphaera watsonii WH0003	✓		GAF
Crocosphaera watsonii WH8501	✓		GAF
Fischerella sp. JSC-11	✓		GAF
Nostoc punctiforme PCC 73102	✓		GAF

Species	Domain architecture of Hik2		
	Full-length	Truncated	Sensor domain prediction
Nostoc sp. PCC 7120	✓		GAF
Cyanobacterium UCYN-A	✓		GAF
Cylindrospermopsis raciborskii CS-505	✓		GAF
Microcoleus chthonoplastes PCC 7420	✓		GAF
Microcoleus vaginatus FGP-2	✓		GAF
Microcystis aeruginosa NIES-843	✓		GAF
Raphidiopsis brookii D9	✓		GAF
Moorea product 3L	✓		GAF
Nodularia spumigena CCY9414	✓		GAF
Anabaena variabilis ATCC 29413	✓		GAF
Nostoc azollae 0708	✓		GAF
Arthrospira maxima CS-328	✓		GAF (not shown)
Arthrospira sp. PCC8005	✓		GAF
Arthrospira platensis str. Paraca	✓		GAF
Lyngbya sp. PCC 8106	✓		GAF
Oscillatoria sp. PCC 6506	✓		GAF
Thermosynechococcus elongatus BP-1	✓		GAF
Trichodesmium erythraeum IMS101	✓		GAF
Acaryochloris marina MBIC 11017	✓		GAF
Synechococcus sp. PCC7335	✓		GAF
Acaryochloris sp. CCMEE 5410	✓		GAF (not shown)
Synechococcus elongatus PCC 7942	✓		GAF
Synechococcus elongatus PCC 6301	✓		GAF (not shown)
Synechococcus sp. PCC 7002	✓		GAF
Synechococcus sp. WH 5701	✓		No prediction
Synechococcus sp. CB 0205	✓		No prediction
Synechococcus sp. RCC 307	✓		No prediction
Synechococcus sp. WH 8016	✓		GAF (not shown)
Synechococcus sp. WH7805	✓		No prediction
Synechococcus sp. CC9605	✓		No prediction
Synechococcus sp. WH 7803	✓		GAF (not shown)
Prochlorococcus marinus str. MIT 9301	✓		No prediction
Prochlorococcus marinus str. AS 9601	✓		No prediction
Synechococcus sp. JA-2-3B'a (2-13)		✓	
Synechococcus sp. JA-3-3Ab		✓	
Gloeobacter violaceus PCC 7421		✓	

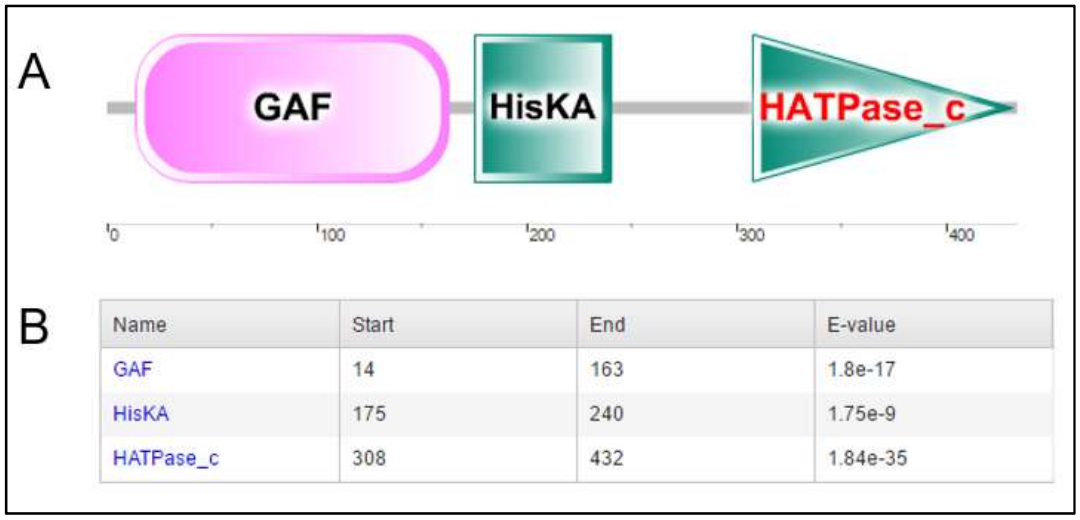
Species	Domain architecture of Hik2		
	Full-length	Truncated	Sensor domain prediction
Cyanidioschyzon merolae	✓		GAF
Phaeodactylum tricornutum	✓		GAF (not shown)
Thalassiosira pseudonana	✓		No prediction
Oryza sativa	✓		No prediction
Arabidopsis thaliana	✓		GAF (not shown)
Populus trichocarpa	✓		No prediction
Physcomitrella patens	✓		GAF (not shown)
Ostreococcus tauri	✓		GAF (not shown)
Ostreococcus lucimarinus	✓		GAF (not shown)
Chlorella vulgaris C-169	✓		No prediction

Table 1.1 Distribution of full-length and truncated forms of Hik2. Tick (✓) indicates the presence of Hik2. Full-length Hik2 proteins (Class I) contain a sensor domain and a kinase domain. Truncated Hik2 proteins (Class II) only have a kinase domain. Chloroplast-containing species are coloured green. 'GAF (not shown)' means there is a possible GAF domain in the sensor domain, but it was not shown in the SMART diagram because the E-value score was less significant than the required threshold. 'No prediction' means the sensor domain could not be predicted to be any known domain in the SMART database. Table modified from (Ibrahim et al. 2017).

1.2.2 Domain architecture of *Synechocystis* Hik2

Synechocystis sp. PCC 6803 is one of the most well studied types of cyanobacteria, and widely used as a model organism. Its whole genome sequence has been determined in 1996 as the first genome of photosynthetic organism (Kaneko et al. 1996). The amino acid sequence of *Synechocystis* sp. PCC 6803 Hik2 (Locus ID: slr1147, GenBank: BAK49478.1, 434 amino acids) was sent to TMHMM (TransMembrane prediction using Hidden Markov Models) Server v. 2.0 (Krogh et al. 2001) to perform the prediction of transmembrane helices, and there is no transmembrane helix predicted in Hik2.

Sequence analysis on SMART online server (Schultz et al. 1998, Letunic et al. 2015) shows that *Synechocystis* Hik2 is a typical histidine kinase, containing a GAF domain as its sensor domain, and two conserved subdomains, a HisKA (the abbreviation of His Kinase A (phosphoacceptor)) domain as its DHp domain and a HATPase_c (Histidine kinase-like ATPase, C-terminal) domain as its CA domain (Figure 1.14).



Name	Start	End	E-value
GAF	14	163	1.8e-17
HisKA	175	240	1.75e-9
HATPase_c	308	432	1.84e-35

Figure 1.14 Domain prediction of *Synechocystis* Hik2. A) Domain architecture prediction. B) The regions and E-values of predicted domains. The prediction was carried out using the SMART online server for identifying possible signalling domains (Schultz et al. 1998, Letunic et al. 2015).

1.2.3 Protein-protein association studies of Hik2

Conducting a protein-protein interaction study helps investigate the intracellular signalling pathways and obtain information of various biochemical processes. The STRING database (Search Tool for the Retrieval of Interacting Genes/Proteins) (Szklarczyk et al. 2015), which contains both known and predicted protein-protein interactions, is used to study the functional associations between Hik2 and other proteins (Figure 1.15). Hik2 is a typical histidine kinase, containing the conserved

phosphoacceptor histidine residue at sequence position 185. It is predictable and has been proved by yeast two-hybrid analysis experiment that Hik2 interacts with its response regulators, as Rres in cyanobacterial, such as Rre1 in *Synechocystis* sp. PCC 6803 (Sato et al. 2007). The rest of the Rres, Hiks, and response regulator RpaA, are mostly predicted to associate with Hik2, as probed by text mining, gene fusion, and gene co-occurrence (Figure 1.15).

The Rre gene products in Figure 1.15 control many stress-related genes, including those involved in temperature and salt tolerance, and they may also regulate genes encoding for the photosynthetic light-harvesting antenna (Los et al. 2010), while RpaA regulates phycobilisome association (Ashby and Mullineaux 1999) and the expression of circadian clock genes (Markson et al. 2013). This is likely to indicate that Hik2 is a multifunctional histidine kinase.

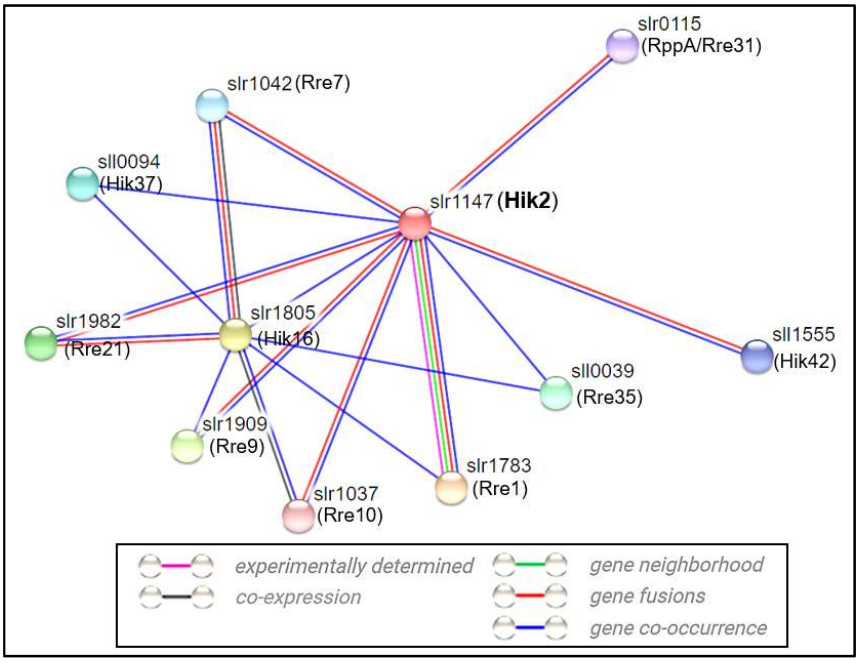


Figure 1.15 Known and predicted protein-protein associations between Hik2 and other proteins from the STRING database (Szklarczyk et al. 2015). Different types of interactions are drawn in straight lines with different colours as listed in the figure.

1.2.4 Function of Hik2

Recent studies proved that Hik2 directly binds not only to Rre1, but also to RppA, and transfers phosphoryl groups to them (Sato et al. 2007, Ibrahim et al. 2016). Rre1 is well conserved in all sequenced cyanobacterial strains (Ashby and Houmard 2006), and has been shown to regulate the transcription of *adhA* gene (coding for alcohol dehydrogenase) *in vitro*, responding to heat and hyperosmotic stress in *Synechocystis* (Vidal et al. 2009). The homologue of Rre1 in non-green algae, Ycf29, has been suggested to bind genes coding for the light-harvesting phycobilisome proteins in response to light intensity decrease (Minoda and Tanaka 2005). In the meantime, RppA in *Synechocystis* has been proved to be involved in the suppression of PS II genes transcription and the increase of PS I genes transcription, in order to regulate the stoichiometry of the two photosystems; and it can also inactivate the destruction of phycobilisome and increase the transcription of phycobilisome genes. Moreover, it has been show that RppA can respond to the redox state changes of the PQ pool (Li and Sherman 2000). Thus, it may be inferred that Hik2 plays an important part in regulating the transcription of photosynthetic genes via these two response regulators.

Figure 1.16, a model of Hik2-based transcriptional control was proposed, for the suggested response regulators Rre1 and RppA regulated by a Hik2 transduction system, in this case for the cyanobacterium *Synechocystis* sp. PCC 6803 (Ibrahim et al. 2016). Here, Hik2 might sense the oxidised PQ and becomes autophosphorylated, then transfers the phosphoryl group to the Rre1 and RppA proteins. Phosphorylated RppA cause the activation of genes coding for PS II and PS I (Li and Sherman 2000), which in turn controls the electron transport between these two photosystems. Moreover, the activity of Hik2 can be inhibited by high concentrations of salt, and phosphorylated

Rre1 represses the expression of salt tolerance genes (Paithoonrangsarid et al. 2004, Shoumskaya et al. 2005). Thus, under high salt conditions, Hik2 cannot phosphorylate Rre1 and the cyanobacterium can respond properly and efficiently to hyperosmotic stress.

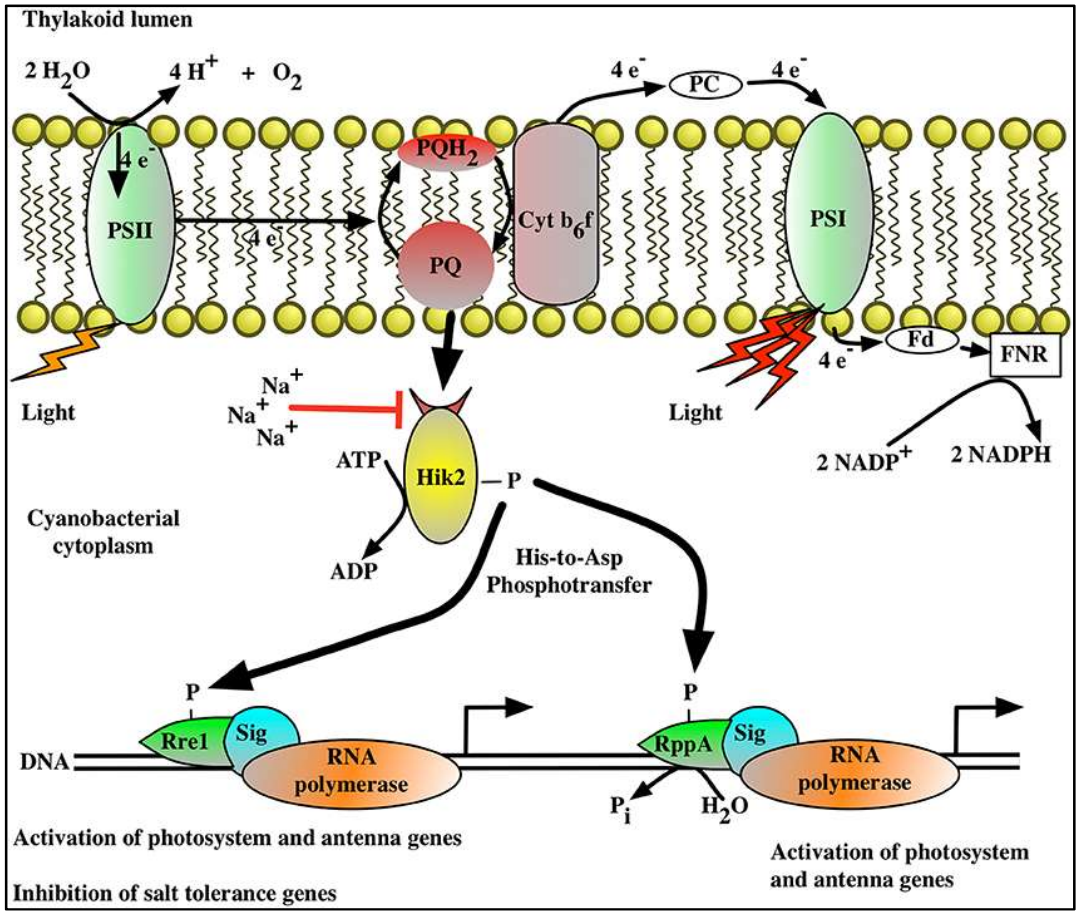


Figure 1.16 Transcriptional control of cyanobacterial genes by Hik2-Rre1/RppA system. The autophosphorylation activity of Hik2 is regulated by signals from photosynthesis and by sodium ions (Na^+). Autophosphorylated (active) Hik2 transfers the phosphoryl groups to Rre1 and RppA. Phosphorylated Rre1 and RppA activate genes encoding for photosynthetic reaction centres. In addition, phosphorylated Rre1 acts as a negative regulator for salt tolerance genes. Under high salt conditions, Na^+ inhibits the autophosphorylation of Hik2, thus removing any repression of the salt tolerance genes. Figure from (Ibrahim et al. 2016).

1.3 Chloroplast Sensor Kinase

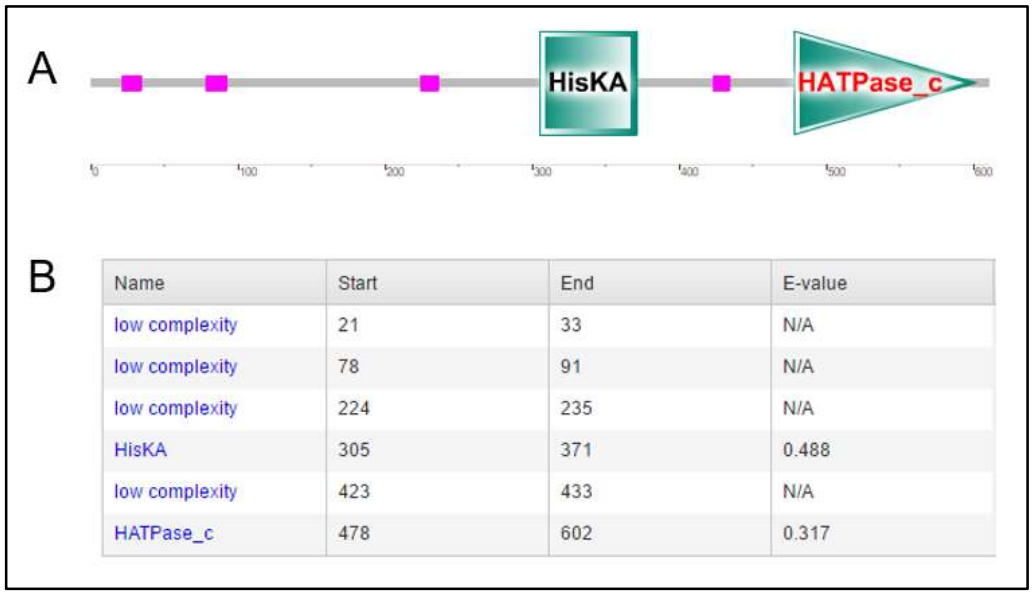
Chloroplast Sensor Kinase (CSK) is now recognised as a homologue of the original sensor histidine kinase Hik2 (Puthiyaveetil et al. 2008), as first found in the model plant *Arabidopsis thaliana* chloroplast, among the 54 proteins that were identified to be related to TCS by sequence analysis, including histidine kinases, histidine-containing phosphotransfer proteins and response regulators (Hwang et al. 2002).

As said, CSK was identified as a bacterial-type sensor kinase in *Arabidopsis*, being the genomic sequence *At1g67840* gene product codes for 611 amino acids. It aligns with sequences of three bacterial histidine sensor kinases and of homologous proteins of other plants and algal species. *Arabidopsis* CSK contains a Glu residue in the homologous H-box motif in DHp domain, instead of the conserved His residue (Puthiyaveetil et al. 2008). In the same 2008 studies, however, the conserved motifs N, G1, F, G2, and G3 in the CA domain can be found in *Arabidopsis* CSK. CSK was observed to be synthesised in cytoplasm and then imported into the chloroplast stroma in precursor form, and the transit peptide was found to be retained in the mature protein complex.

1.3.1 Domain architecture of CSK

The amino acid sequence of CSK (GenBank: AEE34705.1, 611 amino acids) was sent to the TMHMM Server v. 2.0 (Krogh et al. 2001) to perform the prediction of transmembrane helices, and the result shows that there is no transmembrane helix predicted in CSK.

When analysing the *Arabidopsis* CSK sequence with the online software that identifies the protein signalling domains, SMART (Schultz et al. 1998, Letunic et al. 2015), it was shown that no known structure was present for its sensor domain, but present were two conserved histidine kinase domains, similar to Hik2, a HisKA domain and a HATPase_c domain (Figure 1.17 A). The HisKA domain which is a common type of DHp domains, dimerised to form a two-helix bundle. In this domain, there is a conserved phosphor-acceptor His residue that is the site of autophosphorylation in the cyanobacterial homologue of HiK2 under study here, but the typical histidine phosphorylation site has been forced to be replaced by a Tyr residue in the green alga *Ostreococcus taurii*, and replaced by a Glu residue in CSK in land plants including *Arabidopsis thaliana* (Puthiyaveetil et al. 2008), therefore this makes CSK in *Arabidopsis* a modified histidine kinase.



Name	Start	End	E-value
low complexity	21	33	N/A
low complexity	78	91	N/A
low complexity	224	235	N/A
HisKA	305	371	0.488
low complexity	423	433	N/A
HATPase_c	478	602	0.317

Figure 1.17 Domain prediction of *Arabidopsis* CSK. A) Domain architecture prediction. B) The regions and E-values of predicted domains. The prediction was carried out using the SMART online server for identifying possible signalling domains (Schultz et al. 1998, Letunic et al. 2015).

1.3.2 Protein-protein association studies of CSK

The STRING database was used to study the functional associations between CSK and other proteins (Szklarczyk et al. 2015) (Figure 1.18). Histidine kinases usually interact with its cognate response regulators. However, in *Arabidopsis* and other plants, CSK is a modified histidine kinase, and it was found that it does not have a corresponding response regulator (Puthiyaveetil et al. 2008).

As is seen in Figure 1.18, most proteins listed were obtained from co-expression experiments. It has been proved that CSK directly interacts with SIG1 and CKA2 by yeast two-hybrid assay (Puthiyaveetil et al. 2010) (BioGRID interaction database (Stark et al. 2006)). SIG1 is RNA polymerase sigma factor 1, which is needed for initiation of chloroplast RNA synthesis, controlling the transcription of *psaA* gene and thus regulates photosystem stoichiometry (Kanamaru et al. 1999, Shimizu et al. 2010). CKA2 is Casein Kinase 2, regulating plant growth and responding to certain environmental stimuli such as abscisic acid and salt stress (Mulekar and Huq 2014). Other proteins that co-expressed with CSK includes LPA1 (Low PS II Accumulation 1), a chaperone required for efficient PS II assembly (Zapata et al. 2016), PGRL1A (PGR5-like protein 1A), the ferredoxin-plastoquinone reductase involved in cyclic electron flow around PS I (Terashima et al. 2012), and ARASP, a serine protease that is essential for chloroplast and plant development (Zapata et al. 2016), amongst others. The predicted protein-protein association indicates that CSK may function in the photosynthesis process, which is consistent with the literature discussed in the following section.

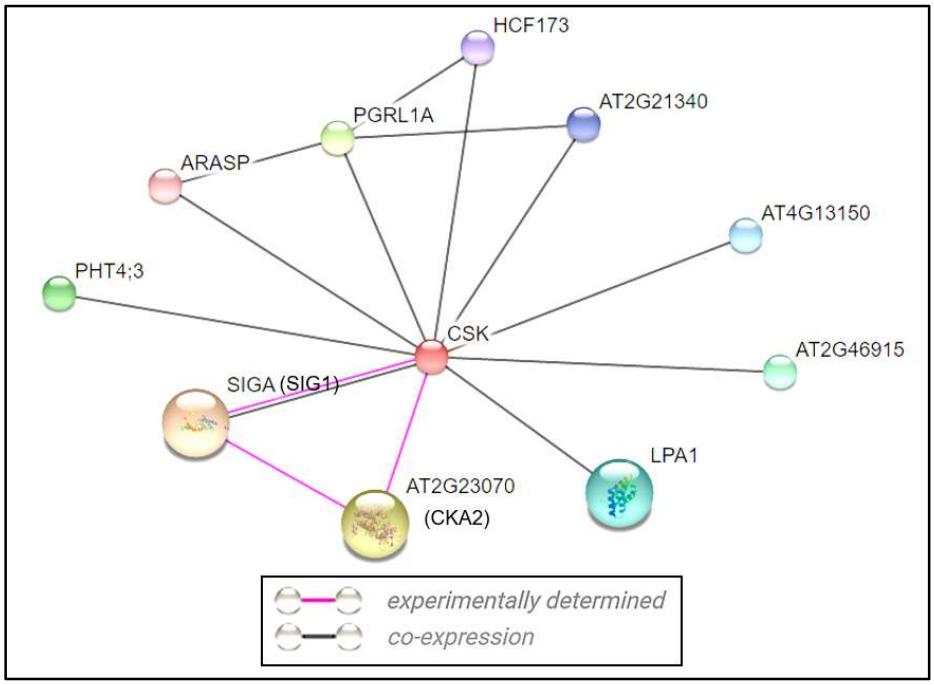


Figure 1.18 Known and predicted protein-protein associations between CSK and other proteins from the STRING database. Different types of interactions are drawn in straight lines with different colours as listed in the figure.

1.3.3 Function of CSK with relevance to Hik2

To date, CSK is known to play a key role in connecting photosynthetic electron transport with chloroplast gene transcription. It achieves this by communicating the redox state of PQ pool to the chloroplast transcriptional apparatus, initiating the appropriate alteration in photosystem stoichiometry (Puthiyaveetil et al. 2008). There is evidence that proves PQ can bind to CSK (Ibrahim et al. 2016), and that CSK interacts with itself, with the bacterial-type RNA polymerase Sigma factor-1 (SIG1), and also with the Plastid Transcription Kinase (PTK) (Puthiyaveetil et al. 2010). SIG1 is required for repression of the *psaA* gene, a gene which encodes a reaction centre apoprotein of PS I, thereby controlling photosystem stoichiometry (Shimizu et al. 2010). In turn, this means it influences directly the rate of photosynthesis as a regulatory

outcome. PTK can phosphorylate Plastid-encoded RNA Polymerase (PEP) and SIG1, resulting in a nonspecific suppression of transcription of all chloroplast genes (Baginsky et al. 1999, Ogrzewalla et al. 2002, Puthiyaveetil et al. 2010). Once again, PEP phosphatase is an unidentified protein phosphatase involved in dephosphorylating phosphorylated PEP and SIG1.

As is shown in Figure 1.19, Puthiyaveetil et al. (2013) proposed a signal transduction pathway for the adjustment of overall photosystem stoichiometry. Here, CSK revealed itself to be an important factor in sensing light quality and the subsequent control of the ratio of PS II vs PS I gene expression. In turn, the efficiency of photosynthesis output is affected. In terms of relevance to the above, when considering Hik2 and its recent biochemical characterisation, which involved investigating its response regulators Rre1 and RppA (Ibrahim et al. 2016) (see section 1.2.4), Hik2 was demonstrated to autophosphorylate and then transfer this phosphoryl group to both of these regulators. This then poses the question how CSK, and therefore Hik2, achieves these outcomes and can be observed to perform them.

The light reactions of photosynthesis can be accelerated by light with a wavelength composition that favours PS II (PS II light) or PS I (PS I light) (Duysens and Amesz 1962, Myers 1971). PS I light contain red- and far-red enriched illumination at a rate-limiting intensity. Under these conditions, the PQ pool is maintained in its oxidised form (PQ) (Figure 1.19 a). PS II light contains low-intensity illumination at wavelengths below the red drop in quantum yield. Now under these latter conditions, the PQ pool is largely in its reduced form (PQH_2) (Figure 1.19 b). Thus, there seems to

be a redox sensory system formed by CSK, PTK and SIG1 in the regulation of photosystem stoichiometry (Puthiyaveetil et al. 2013).

Under PS I light, CSK senses PQ signal and becomes autophosphorylated, the active form being a protein kinase, which can phosphorylate SIG1 and PTK. Phosphorylated SIG1 represses the transcription of *psa* genes, which encode for PS I reaction centre proteins, but *psb* genes (coding for PS II reaction centre proteins) are not affected. PTK becomes inactive after being phosphorylated by CSK, and will not suppress the transcription of chloroplast genes (Link 2003, Puthiyaveetil et al. 2010). PEP phosphatase is always active, but the rate of dephosphorylation on phosphorylated PEP is exceeded by that of phosphorylation on SIG1. So, under these conditions, the transcription of *psb* genes of PS II surpasses that of *psa* genes for PS I, resulting in an increased and increasing ratio of PS II/PS I.

On the contrary, under PS II light, CSK cannot sense any PQ signal and is not autophosphorylated, remaining in its inactive form as a protein kinase, hence SIG1 and PTK remain dephosphorylated. However, the reduced glutathione concentration rises, inhibiting the activity of PTK (Baginsky et al. 1999, Baena-Gonzalez et al. 2001), so PTK is only partially active. In this environment, the rate of PTK phosphorylating PEP is less than that of PEP phosphatase dephosphorylating PEP, and this then results in a continuous transcription of both *psa* and *psb* genes. Therefore, the increasing transcription rate of *psa* genes leads to an increase in PS I relative to PS II (Figure 1.19 b), and the organism is able to retain an increased ability to photosynthesis under more challenging, non-ideal, light quality.

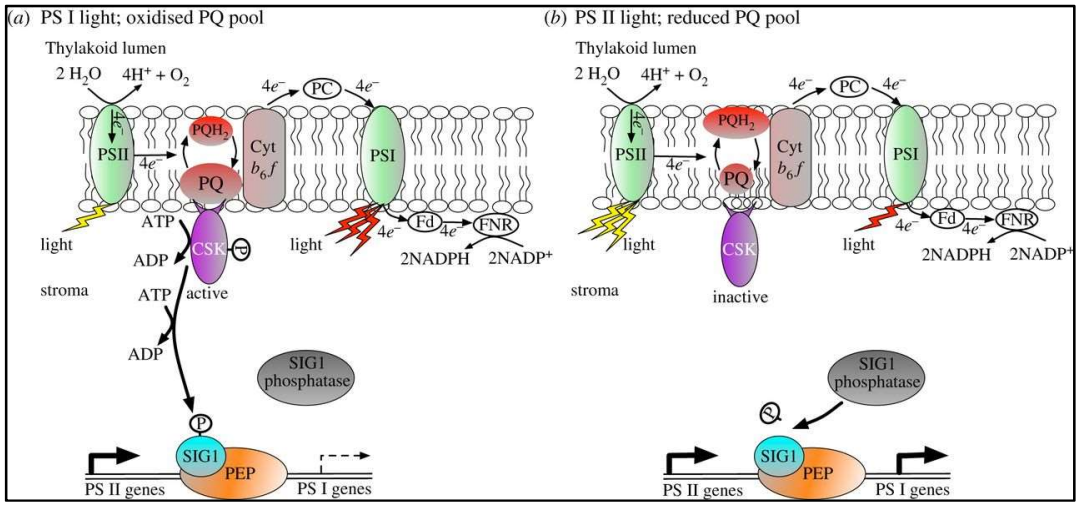


Figure 1.19 A regulatory scheme of CSK for plant and algal photosystem stoichiometric adjustment. a) Under PS I light, CSK becomes autophosphorylated and active, resulting in the repression of PS I genes, but PS II genes are not affected. b) Under PS II light, CSK is not autophosphorylated and inactive, and the SIG1 phosphatase makes phospho-SIG1 dephosphorylated, thus removing the repression of PS I gene transcription. Figure from (Puthiyaveetil et al. 2013).

1.4 Photosynthesis

As introduced in the last two sections, both Hik2 and CSK have been suggested to play a key role in photosynthesis, one of the most fascinating biochemical processes on Earth. In the Greek language, the word ‘photosynthesis’ means ‘building up from light’, and since 1893, Charles Barnes proposed this word to replace the old name ‘assimilation’ (Gest 2002). It is a chemical process through which plants, and a variety of other organisms such as algae and many species of bacteria, are able to capture and convert solar energy into chemical energy, in order to maintain their life-dependant activities. Energy from our primary source, the sun, is absorbed by protein complexes named photosynthetic reaction centres belonging to two photosystems, Photosystem II (PS II) and Photosystem I (PS I), each being large multi-subunit supercomplexes

defined by the *psb* and *psa* gene products, respectively, linked together as defined by the redox-based 'Z scheme' (Govindjee 1975). In plants, they are held inside organelles called chloroplasts, while in bacteria the reaction centres are embedded in the thylakoid membrane (Blankenship 2014).

In green plants, chlorophyll-type pigments are used to capture light by the most typical derivative of photosynthesis, the oxygenic form, driven at its core by the light-induced electron transfer process termed the photosynthetic electron transport chain. This photosynthetic process is housed in the chloroplasts of green plants, where thylakoid membrane-bound proteins termed the Light Harvesting Complex (LHC) absorb photons and transfer the energy to a first photosystem in the chain, PS II, leading to an oxidation of a water molecule. There are two major families of LHC proteins; those that service the PS II core, termed LHC II, and those that service the other photosystem, PS I, which belong to LHC I. The electrons extracted from water are then transferred to a sequence of specific PS II cofactors, those being chlorophylls, pheophytin and quinones, thereby, in effect, spatially crossing the thylakoid membrane. Plastoquinone (PQ), an important mobile electron carrier present in the thylakoid membrane, is firstly reduced by accepting two protons from the stromal matrix of the chloroplast and two electrons from PS II, becoming the reduced form Plastoquinol (PQH_2). PQH_2 transports the protons to the inner space, or lumen, of the thylakoid, while the electrons pass through the photosynthetic electron transport chain into the dimeric Cytochrome (Cyt) b_6f protein complex, then finally to a second photosystem, PS I, via an extrinsic mobile electron carrier, Plastocyanin (PC), of the lumen. The electrons are then transferred through Ferredoxin (FD) and its flavin cofactor Flavin Adenine Dinucleotide (FAD), eventually to an intermediate electron acceptor $NADP^+$, reducing it to NADPH. The H^+ produced

from the entire photosynthetic electron transport chain is pumped into the thylakoid lumen in order to generate a H^+ gradient between the stroma and lumen, which is used to synthesise ATP when it flows back, through ATPase, ending up finally into the stromal matrix after producing the energy required for cellular processes (Mitchell 1961). Significantly for humanity, both NADPH and ATP can be utilised to reduce carbon dioxide (CO_2) to sugars, ultimately providing for our food and may eventually mitigate the effects of climate change (Figure 1.20).

Looking specifically at the mobile electron carrier, PQ, its reduction-oxidation (redox) state, which is measured by the ratio of PQH_2 to PQ, controls the reversible phosphorylation of a number of proteins including LHC II. LHC II is redistributed between the two photosynthetic reaction centres, and it has been shown that the ratio of PQH_2 to PQ governs the distribution of absorbed light energy between PS II and PS I (Allen et al. 1981, Allen 1992). The redox state of PQ also regulates the expression of genes *psbA* and *psaA/psaB*, which encode the D1 protein of PS II and the reaction centre apoprotein A1/A2 of PS I, respectively (Pfannschmidt et al. 1999, Tullberg et al. 2000, Puthiyaveetil and Allen 2008). This results in a long-term gene-specific response to light quality now known as photosystem stoichiometric adjustment (Pfannschmidt et al. 1999).

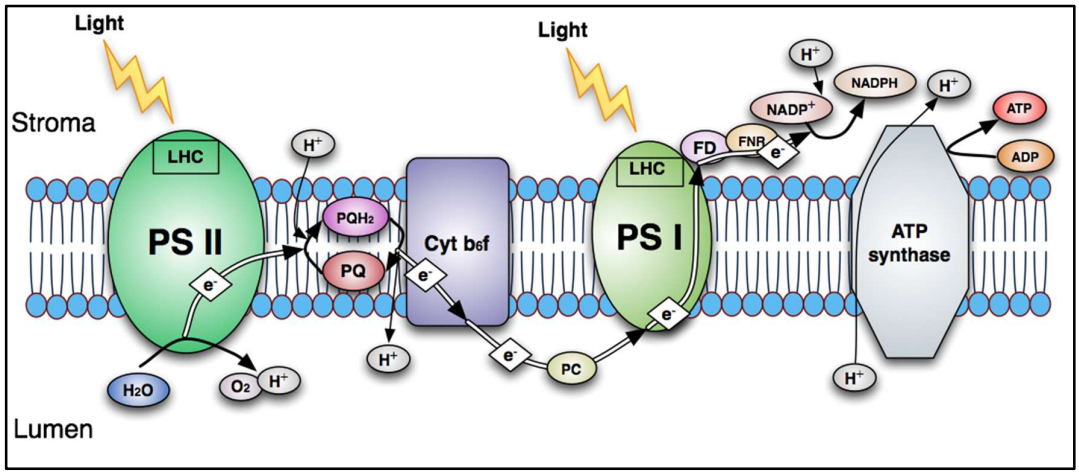


Figure 1.20 Schematic model of the photosynthetic electron transport chain within a typical thylakoid membrane of the chloroplast. Model is exemplified on the basis of the known literature for *Arabidopsis thaliana*. PS II and PS I refer to two functional units in photosynthesis, photosystem II and photosystem I; LHC is the light harvesting complex; Cyt b_6f is cytochrome b_6f complex, the plastoquinol-plastocyanin reductase; PQ (plastoquinone), PQH_2 (plastoquinol), PC (plastocyanin) and FD (ferredoxin), are mobile electron carriers; FNR is the ferredoxin-$NADP^+$ reductase; ATP (adenosine 5'-triphosphate) and NADPH (nicotinamide adenine dinucleotide phosphate) are high energy compounds that drive the chemical reactions of sugar synthesis; $NADP^+$ is the oxidised form of NADPH; ADP is adenosine 5'-diphosphate.

Apart from green plants, photosynthesis is also performed in algae and cyanobacteria. According to the endosymbiotic theory, cyanobacteria are the ancestors of plant chloroplasts, this being supported by comparisons made between their genomes and intrinsic functions (Martin and Kowallik 1999, Deusch et al. 2008). There are similarities and also differences between these photosynthesis processes occurring in cyanobacteria and chloroplasts. Photosynthesis in most cyanobacteria takes place only within the thylakoid membranes, but not in the cytoplasmic membranes, while the respiration takes place in both (Gantt 1994, Schmetterer et al. 1994), hence the photosynthetic and respiratory processes share some electron transfer complexes,

including the PQ, the Cyt b_6f complex, and various soluble electron carriers in the lumen, such as PC. However, the electron transport chain of photosynthesis in cyanobacteria remains similar to that of plants. Figure 1.21 is a schematic model of the photosynthetic electron transport chain in the cyanobacterial thylakoid membrane, showing the protein complexes engaged in both photosynthesis and respiration, as based on the current literature for *Synechocystis* sp. PCC 6803 (Liu 2016). Here, during photosynthesis, the phycobilisome (sometimes termed the phycobilisome mass) (Glazer 1985) serves as extrinsic (to the membrane) light antenna to capture light energy, and rapidly transfers this energy to PS II, in which the light-induced electron transfer reactions happen, bringing about the water splitting reaction and then PQ pool reduction. Reduced PQ is next able to transfer electrons to PS I via Cyt b_6f and PC. The FD is subsequently reduced and transfers electrons to $NADP^+$, forming NADPH, which is used for CO_2 fixation and glucose production.

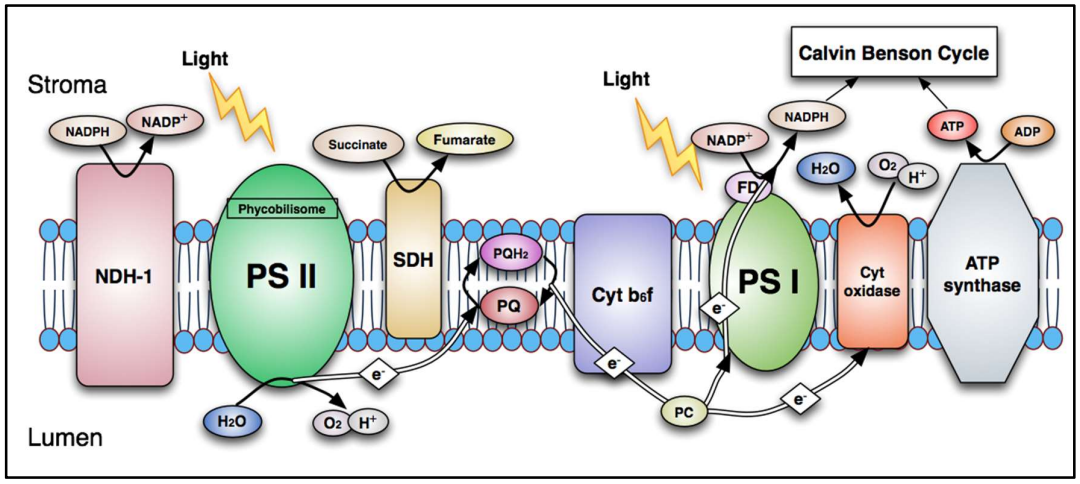


Figure 1.21 Schematic model of photosynthesis and respiration in the cyanobacterial thylakoid membrane. Model is exemplified based upon the known knowledge for *Synechocystis* sp. PCC 6803. NDH-1 is the quinone oxidoreductase; SDH, the succinate dehydrogenase; Cyt oxidase, the cytochrome oxidase. The remaining abbreviations are as per the figure legend of Figure 1.20.

The importance of photosynthesis, directly bearing on the evolution that has culminated in humanity, is abundantly clear. The splitting of water by PS II into protons (H^+), a source of renewable fuel, and oxygen, the gas that animals require to respire, coupled to the capture of CO_2, in the Calvin-Benson cycle, providing for life's primary food source, renders Hik2 and the kinase family, in general, a role that is not insignificant. As described, in photosynthesis, CSK connects the redox state of the PQ pool to chloroplast reaction centre genes transcription, through the phosphorylation of SIG1 and PTK, initiating an appropriate alteration in photosystem stoichiometry (Puthiyaveetil et al. 2013). In the case of Hik2, it has been proved to phosphorylate its cognate response regulators Rre1 and RppA, activating the genes coding for PS II and PS I, this being in response to other photosynthesis related redox signals (Ibrahim et al. 2016). Therefore, both CSK and Hik2 are essential for the regulation of photosystem genes transcription.

1.5 Aims and objectives of this work

My PhD research aims to probe the structure of *Synechocystis* Hik2, the cyanobacterial homologue of CSK, in order to elucidate how Hik2 performs its histidine kinase functional role. CSK was previously shown to play an important part as a novel redox sensor in the chloroplasts of green plants, using the model plant *Arabidopsis thaliana*. In my project, I am building upon the knowledge that Hik2 has been shown to control transcription of genes coding for photosystems in cyanobacteria, and Na^+ ions inhibit the autophosphorylation activity. It would be important to explore the mechanism by which the activity of Hik2 is switched off as well as how Hik2 perceives signals.

During my research, a variety of Hik2 structural populations were observed, by size exclusion chromatography, native-PAGE and cross-linking analysis, all observations further validated by negative stain Transmission Electron Microscopy (TEM). In addition, the sensor domains of *Synechocystis* Hik2 (GAF domain) and *Arabidopsis* CSK (ligand binding domain) were analysed, with the goal of revealing their signal sensing mechanisms. In the first results chapter of this thesis, Chapter 3, I describe the oligomeric states of Hik2 and its subdomains, from two cyanobacterial species, and discuss the crucial DHp domain and the dissociation effect of NaCl. In Chapter 4, attempts were made to purify and solubilise the sensor domains of Hik2 and CSK by *in vitro* denaturation and refolding, and ligate them with MBP and SUMO tags. Analysis of limited proteolysis was also applied to Hik2 protein. Since the direct visualisation of a protein can be obtained from TEM image processing, combined with computer-based single particle image analysis, such analyses were performed and described in Chapter 5, and several novel, characteristic, 2D projection views of Hik2 oligomers were revealed.

Chapter 2 **Materials and methods**

2.1 Plasmid construction and transformation

In this thesis, the full-length *hik2* gene and the DNA sequence of its DHp domain and kinase domain from *Thermosynechococcus elongatus* BP-1 were amplified and cloned into a pET_21b vector. The sensor domain of the *Synechocystis* sp. PCC 6803 *hik2* gene was also amplified, and the sensor domain of the *Arabidopsis thaliana csk* gene was commercially synthesised, and these genes were each cloned into pET_14b, pOPINM, and pOPINS3C vectors. All recombinants were transformed into *E. coli* BL21-(DE3) competent cells in order to overexpress the target proteins.

2.1.1 Polymerase chain reaction

Polymerase Chain Reaction (PCR) was performed to amplify the DNA using a C1000 thermal cycler (Biorad) and Q5® High-Fidelity DNA Polymerase (New England BioLabs). All reactions were carried out in a total volume of 50 µl within thin wall 200 µl PCR tubes (Starlab). A typical PCR reaction composition is shown in Table 2.1, and characteristic PCR thermal cycle program, in Table 2.2.

Component	Volume (µl)	Final concentration
5 × Q5 Reaction Buffer	10	1 ×
10 mM dNTPs	1	200 µM
10 µM Forward Primer	2.5	0.5 µM
10 µM Reverse Primer	2.5	0.5 µM
Template DNA	1	10 ng/µl
Q5 High-Fidelity DNA Polymerase	0.5	0.02 U/µl
Nuclease-Free Water	32.5	
Total	50	

Table 2.1 A typical PCR reaction composition.

Step	Temperature (ºC)	Time (s)
Initial Denaturation	98	30
	98	10
30 Cycles	50 to 72	20
	72	20 to 40
Final Extension	72	120
Hold	4	

Table 2.2 A standard PCR thermal cycle program. The temperature of the annealing step was usually performed at 3 ºC above the lower Tm of the primer Tm. The time of the extension step was generally set up as 20 to 30 s per kb.

The full-length *Thermosynechococcus elongatus* BP-1 (sourced from a hot spring in Beppu, Japan) gene *hik2* (UniProtKB: Q8DMC5) (coding for amino acids 1 to 385), as well as the truncated version (kinase domain) (coding for amino acid 143 to 385) and also the DHp domain (for dimerisation and phosphor-accepting function) (coding for amino acid 143 to 270) were amplified from template DNA, and cloned into pET_21b vectors (with a C-terminal his-tag, Novagen). The GAF domain (sensor domain) (coding for amino acid 14 to 163) of the *Synechocystis* sp. PCC 6803 *hik2* gene (GenBank: BAK49478.1) was amplified from template DNA, and cloned into pET_21b, pOPINM (with an N-terminal His-MBP-3C tag), and pOPINS3C (with an N-terminal His-SUMO-3C tag) vectors. pOPINM was a gift from Ray Owens (Addgene plasmid # 26044). pOPINS3C was a gift from Ray Owens (Addgene plasmid # 41115). *Thermosynechococcus elongatus* BP-1 and the *Synechocystis* sp. PCC 6803 genomic DNA was kindly donated by Dr. Dennis J. Nuernberg, Imperial College London. Primer pairs (purchased from Eurofins MWG Operon, Germany) are listed in Table 2.3.

Primer	Sequence	Restriction site
Thermo_Hik2_F		
Forward	GCGCcatatgATGCTCTGGCCAGCCAGT	NdeI
Reverse	GCGCGCgtcgacTGGTTCCACCTTCATTTG	SalI
Thermo_Hik2_Truncated		
Forward	GCGCcatatgATGCACTCCCCTGCCCAGCCA	NdeI
Reverse	GCGCGCgtcgacTGGTTCCACCTTCATTTG	SalI
Thermo_Hik2_DHp		
Forward	GCGCcatatgATGCACTCCCCTGCCCAGCCA	NdeI
Reverse	GCGCgtcgacTTCCTCGAGCCAGATCGG	SalI
Sync_Hik2_GAF		
Forward	GGcatatgGGGTTAATTTCCCTT	NdeI
Reverse	CCCctcgagTTAGCGATACCAATC	XhoI
CSK_LBD		
Forward	GGcatatgATTTTTACCAGA	NdeI
Reverse	CCCggatccTTACATTGGAA	BamHI

Table 2.3 Primers used for cloning. Restriction site base pairs are written in lower case.

The Ligand Binding Domain (LBD) of the *Arabidopsis thaliana csk* gene (At1G67840, GenBank: AEE34705.1) (coding for amino acid 51 to 207) was commercially synthesised by the company Eurofins Genomics Ltd, Ebensburg, Germany, under the consideration of optimised codon usage in *E. coli* (Puigbo et al. 2007), and cloned into pET_14b (with an N-terminal his-tag, Novagen), pOPINM, and pOPINS3C vectors. The synthesised DNA sequence is listed in Table 2.4.

CSK_LBD
GCGCcatatgATTTTTACCAGAGGACTTCGGTATGTCAACCACACCGTCAGCAA CGAAGAATCCGAGCCTGGTGGCGGCGAGACGATGGTAGCGAGTGCATCGGC TATTGCGAGCGCCATAAGAGGTGCATCAACAACGCCGGTAGAGTTTACCCA AATGATTGAAAAGGACCACTTGAAGACAAAGATTATATTGCCCAGTCCCGA CTTCCAAAGATTATGTTTAGAGCAACTGGACTTATTTAGACAGATCGTCGAT CCCAACGCTGTGTTATCGATTTACGTGAGACCCGCCGGCAGTTATGTTATGG ATCGTCTGGAACTGAGACGGGTTACGTGCTACCCATCAGTAAACGCTGGGG ACGTCGTTATTTTAGTTGGGAACTTTGGCATACCGGCGGGCTTGCGTGCGGC AGAAGCGAGCTTGTCGTCTCAGCAGGTAGCAGCTGGTATCGAAACATCGGG CGGCTGTTTTTCCAATGTAAggatccGCGC

Table 2.4 The synthesised DNA sequence of CSK_LBD. Restriction site base pairs are written in lower case.

2.1.2 Agarose gel electrophoresis

Agarose (Sigma) was dissolved in 1 × TAE buffer to a concentration of 1% (w/v), and the ethidium bromide solution (Fisher) was added to a final concentration of 50 μg/ml. The 1% agarose solution was cast into a tank with a comb (Biorad). When the gel was set, DNA samples were mixed with 6 × DNA loading buffer (Invitrogen) and loaded onto the gel, then subjected to electrophoresis at a constant voltage of 2 V/cm until sufficient separation was achieved. DNA bands in the gel were visualised under an illumination of a UV light platform (Uvitec) and excised from the gel to perform the DNA purification.

2.1.3 DNA purification

Amplified PCR product, or excised agarose gel, were treated with the peqGOLD cycle-pure kit (Peqlab) to remove oligonucleotides and enzymes, according to manufacturer's

instruction. Plasmid DNA was purified from an overnight culture of *E. coli* cells through the use of a peqGOLD Miniprep kit (Peqlab), according to the manufacturer's instructions.

Final DNA concentration was detected by UV spectrophotometer according to the equation:

$$\text{Concentration } (\mu g/ml) = (A_{260} - A_{320}) \times \text{dilution factor} \times 50\ \mu g/ml$$

An A_{260}/A_{280} ratio between 1.7 and 2.0 generally represents a high-quality DNA sample. A lower A_{260}/A_{280} ratio may indicate the presence of protein, phenol or other contaminants that absorb strongly at or near 280 nm (Green and Sambrook 2012).

2.1.4 Restriction endonuclease digestion

Double restriction digestion, for PCR products and plasmid DNA, was carried out under the presence of the appropriate buffer (New England BioLabs) in a total volume of 50 µl. All digestions were carried out at 37 ºC for 1 h. A typical reaction mixture is shown in Table 2.5.

Component	Volume (µl)	Final concentration
10 × Reaction Buffer	5	1 ×
Restriction enzyme 1 (10 U/µl)	2	10 U/µg of DNA
Restriction enzyme 2 (10 U/µl)	2	10 U/µg of DNA
DNA	8	0.04 µg/µl
H_2O	33	
Total	50	

Table 2.5 A typical restriction enzyme digest reaction composition for PCR products and plasmid DNA.

2.1.5 DNA ligation

DNA ligation reactions were carried out using a total reaction volume of 20 µl, with the double digested plasmid and insert DNA in 1: 3 molar ratios. The ligation was incubated at 16 ºC overnight or room temperature for 10 min using T4 DNA Ligase kit (New England BioLabs). A typical reaction mixture is shown in Table 2.6.

Component	Amount
10 × T4 DNA Ligase	2 µl
Vector DNA (4 kb)	50 ng (0.02 pmol)
Insert DNA (1 kb)	37.5 ng (0.06 pmol)
T4 DNA Ligase	1 µl
Nuclease-Free Water	to 20 µl
Total	20 µl

Table 2.6 A typical ligation reaction composition.

2.1.6 Recombinant transformation

For plasmid amplification, 5 µl of any given ligation reaction was transformed into One Shot TOP10 chemical competent cells (Invitrogen); for protein over-expression, 2 ng of plasmid DNA was transformed into BL21-(DE3) chemical competent cells (Invitrogen). All transformations were carried out according to the manufacturer's instructions supplied with the competent cells (Invitrogen). A typical transformation was performed as follows: the applicable amount of DNA was added into 50 µl of competent cells, and incubated on ice for 30 min. The vial was then heat-shocked at 42 ºC for 30 s, and placed on ice for 2 min. 250 µl of pre-warmed SOC medium was added into the competent cells, and incubated in a 37 ºC shaker for 1 h. 20 to 200 µl of the culture was plated on Luria Broth (LB) agar plates containing appropriate antibiotics (ampicillin in this thesis, Sigma). All plates were inverted and incubated overnight at 37 ºC.

2.2 Protein production

2.2.1 Overnight culture

10 ml of LB medium with ampicillin, at a final concentration of 100 μg/ml, was inoculated with a single colony of *E. coli* cells or from a glycerol stock. The culture was incubated overnight using a 37 ºC shaker (New Brunswick) at 220 rpm.

2.2.2 Gene over-expression in a large scale

1 L of LB medium with ampicillin, at a final concentration of 100 μg/ml, was inoculated with an overnight culture at 1/100 dilution, and grown at 37 ºC to an Optical Density (OD), at 600 nm, of 0.5 to 0.6. Gene over-expression was initiated with adding the Isopropyl β-D-1-thiogalactopyranoside (IPTG) (Apollo Scientific) to a final concentration of 0.8 mM, and the culture grew for a further 16 h at 18 ºC.

2.2.3 Cell disruption and lysate clarification

Induced cells that contained target proteins were harvested by centrifugation at 6,000 g (Beckman Coulter) for 20 min at 4 ºC, and the pellet was resuspended in the binding buffer (20 mM Tris-HCl, 300 mM NaCl, 25 mM imidazole (pH 7.5) for proteins with his-tag using Ni^{2+} affinity chromatography, or 20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA, 1 mM DTT (pH 7.5) for proteins with MBP-tag using amylose affinity chromatography). Lysozyme (Novagen) (to a final concentration at 300 μg/ml) and DNase (New England BioLabs) (to a final concentration at 1 μg/ml) were added to the cell suspension, and incubated at 4 ºC for 4 h to aid lysis and reduce lysate viscosity. The cell suspension was then applied to an EmulsiFlex-C3 homogenizer (Avestin) or a

sonicator (Bandelin), according to the manufacturer's instructions. Lysate was separated by centrifugation at 18,000 g (Beckman Coulter) for 20 min at 4 ºC.

2.3 Protein purification

2.3.1 Ni^{2+} affinity chromatography

The recombinant proteins with hexahistidine tag (6 his-tag) were purified by Ni^{2+} affinity chromatography as follows, according to the manufacturer's instructions. The soluble fraction after cell disruption was applied to the binding buffer pre-equilibrated Ni^{2+}-charged sepharose column (GE Healthcare), and then washed with 20 Column Volumes (CV) of washing buffer I (20 mM Tris-HCl, 300 mM NaCl, 50 mM imidazole (pH 7.5)) and 2 CV of washing buffer II (20 mM Tris-HCl, 300 mM NaCl, 100 mM imidazole (pH 7.5)) to eliminate non-specifically bound protein. The target protein was eluted with elution buffer (20 mM Tris-HCl, 300 mM NaCl, 500 mM imidazole (pH 7.5)) until no further protein came through (as tested by Bradford solution).

2.3.2 Amylose affinity chromatography

The recombinant proteins with the Maltose-Binding Protein tag (MBP-tag) were purified by amylose affinity chromatography as follows, once again, as according to the manufacturer's instructions. The soluble fraction after cell disruption was diluted 5 times with the binding buffer (20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA, 1 mM DTT (pH 7.5)), and applied to the binding buffer pre-equilibrated amylose resin column (New England BioLabs), then washed with 20 CV of binding buffer. The target protein was eluted with elution buffer (20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA, 1 mM

DTT, 10 mM maltose (pH 7.5)) until no further protein came through (as confirmed by Bradford solution).

2.3.3 Protein concentration determination

Protein concentration was quantified by UV-visible absorption spectroscopy, using a UV-Vis spectrophotometer (Hitachi). Protein absorbance at 280 nm was measured. Concentration was calculated according to the Beer-Lambert's law equation:

$$A_{280} = \varepsilon \times c \times l$$

Where A_{280} is the absorbance at 280 nm, ε is the molar extinction coefficient in $M^{-1}cm^{-1}$ (calculated using ExPASy ProtPrama), c is the molar concentration in M, and l is the path length of the cuvette in cm.

2.3.4 Size exclusion chromatography

Size exclusion chromatography was used to separate large molecules or macromolecular complexes, such as proteins, by their size, and in some cases molecular mass. The eluted protein samples from Ni^{2+} affinity chromatography or amylose affinity chromatography were loaded onto a Superdex 200 HR 10/30 column (GE Healthcare) using an ÄKTA purifier system (GE Healthcare). The column was pre-equilibrated with 2 CV of size exclusion chromatography running buffer and run with 1.5 CV of running buffer at 0.5 ml/min with 1 ml fractions collected. Fractions were checked using SDS-PAGE or native-PAGE. Running buffer is either low salt buffer (20 mM Tris-HCl, 10 mM NaCl (pH 7.5)) or high salt buffer (20 mM Tris-HCl, 500 mM NaCl (pH 7.5)).

2.4 Biochemical and biophysical analysis

A variety of experiments described in this section were performed with the full-length Hik2 protein and its subdomain protein complexes, in order to probe their structural characterisation.

2.4.1 SDS-PAGE

Sodium Dodecyl Sulfate (SDS) Polyacrylamide Gel Electrophoresis (PAGE) was used to separate proteins according to their masses. SDS is an anionic detergent, and binds to a polypeptide chain in proportion to its relative molecular mass. Most of the protein structure is destroyed by the negative charges of SDS and becomes linear, being then strongly attracted toward an anode in an electric field (Ninfa et al. 2009). Thus, the protein mass/charge ratio is almost constant, resulting in the mass of a polypeptide being the dominant factor in determining the rate of migration through the gel.

Gel cast assembly was set up according to the manufacturer's instructions (Bio-Rad). The composition of the gels is given in Table 2.7. Tetramethylethylenediamine (TEMED, Thermo Scientific) was added to the mixture just before pouring it into the cast, in order to catalyse the acrylamide polymerisation. The resolving gel was poured into the cast first, and followed by the stacking gel. A 10-wells or 15-wells comb was inserted before the gel was left to set.

Solution (ml)	4% Stacking (ml)	10% Resolving (ml)	12% Resolving (ml)
30% acrylamide	0.255	1.67	2
1.5 M Tris (pH 8.8)	-	1.25	1.25
1.0 M Tris (pH 6.8)	0.1875	-	-
10% SDS	0.015	0.05	0.05
10% ammonium persulfate	0.015	0.05	0.05
H_2O	1.02	1.98	1.65
TEMED	0.0015	0.002	0.002
Total	1.5	5	5

Table 2.7 Recipe for the preparation of reducing SDS-PAGE gels.

Protein samples were mixed with SDS-PAGE loading buffer (50 mM Tris-HCl, 100 mM DTT, 2% SDS, 0.1% bromophenol blue, 10% glycerol) and heated at 95 ºC for 5 min, then loaded onto each gel. Gels were run in SDS Tris-Glycine running buffer (25 mM Tris-HCl, 192 mM Glycine, 0.05% SDS) at 150 V for 1.5 h or until the desired separation was reached. The gel was stained in Coomassie blue dye solution (0.1% Coomassie Blue R-250 (Sigma), 50% ethanol, 7% acetic acid) for 10 min on a shaker and destained in destaining buffer (10% ethanol, 7% acetic acid) until the protein bands were clearly observable.

For normal SDS-PAGE, a reducing agent, such as Dithiothreitol (DTT, New England Biolabs) or 2-Mercaptoethanol (2-ME, Sigma), was included in the sample loading buffer and running buffer, thus breaking intra/inter-molecular disulfide bonds. For non-reducing SDS-PAGE, reducing agents were omitted from the sample loading buffer and running buffer, and the samples not heated, in order to retain disulfide bonds for any chemical cross-linking analysis.

2.4.2 Native-PAGE

Native (non-denaturing) PAGE is run in the absence of SDS and reducing agents. In native-PAGE, the protein mobility depends on the protein charge/mass ratio, the physical shape, and its size. Since Coomassie Brilliant Blue is a dye that is used in blue native-PAGE and may cause protein complexes to dissociate, clear native-PAGE was used in these studies. To exclude the effect of any protein charge difference, a gradient native gel was also applied to probe the molecular mass of the protein complexes. In so doing, the preservation of protein quaternary structure allowed for the observation of protein oligomerisation under conditions close to those of the *in vivo* environment (native or near-native state).

A native-PAGE gel was cast using the same protocol as for an SDS-PAGE gel (section 2.4.1) except for the lack of SDS and the presence of 10% glycerol in the resolving layer (Table 2.7). 4% to 20% gradient precast native gels were purchased from Bio-Rad Laboratories. Both native gels were run at 4 ºC in Detergent-free running buffer (50 mM Tris-HCl, 400 mM Glycine, 1 mM EDTA) at 120 V for 2.5 h or until the desired separation was reached. Staining and destaining for native-PAGE gels were run as per the same protocol as for an SDS-PAGE gel (section 2.4.1.1).

2.4.3 Limited proteolysis

Limited proteolysis is a biochemical method for probing protein conformational features. When the target protein is incubated with a relatively low concentration of protease, such as trypsin, the protease will cut at recognition sites (Lysine (Lys) and Arginine (Arg) residues for trypsin) throughout the protein, normally at exposed regions

such as loops and other flexible regions. This will provide information regarding the secondary and tertiary structure.

The purified protein was buffer exchanged into 20 mM Tris-HCl, 150 mM NaCl (pH 7.5) using PD-10 desalting column (Amersham Biosciences) or by size exclusion chromatography, and incubated for 24 h at 22 ºC or 4 ºC with trypsin (Sigma) at a protein: trypsin molecular ratio as 100: 1 or 1000: 1. Sample aliquots were taken at various time points, and the reaction terminated with 0.5 mM Phenylmethylsulphonyl Fluoride (PMSF). Samples were analysed by SDS-PAGE or native-PAGE. The band corresponding to the stable fragment obtained after 1 h incubation at a protein: trypsin ratio as 1000: 1 was excised from the SDS-PAGE gel, incubated in 10 µl of distilled water, and sent to the Protein & Nucleic Acid Chemistry Facility at the University of Cambridge for N-terminal sequencing analysis.

2.4.4 Dynamic light scattering

Dynamic Light Scattering (DLS) was used to determine the size distribution profile of small particles in suspension, or polymers in solution. DLS experiments were carried out with a DynaPro Molecular sizing instrument (Protein Solutions^{TM}) controlled by the DYNAMICS V6 software. 45 µl quartz cuvettes were used to hold the sample during measurements. The quartz cuvettes were prewashed with 1% Triton X-100 and subsequently washed with water and ethanol, and dried with compressed air. The exterior surface of cuvette was wiped with ethanol to remove any dirt. For each sample, a minimum of 20 measurements were recorded at 4 ºC.

2.4.5 Circular dichroism spectroscopy

Circular Dichroism (CD) spectroscopy was used to collect the secondary structure information of the full-length *Synechocystis* Hik2 protein. The protein samples prepared for CD were exchanged to buffer containing 20 mM Tris-HCl, 10 mM NaCl (pH 7.5), and diluted with the same buffer to a concentration of 0.3 mg/ml, this being determined using the molar extinction coefficient. For the analysis that Hik2 binds to ATP, 0.1 mM of ATP was added to both the buffer and the protein solution. Far-UV CD measurements were performed using a Jasco J-715 spectropolarimeter equipped with a PTC-348WI temperature controller. The 1 mm path length fused silica cuvette holding the protein sample was constantly orientated in the same direction, in the sample chamber of the CD spectrometer, in order to minimize the variation being introduced into any experiment. Three independent measurements were carried out and the delta absorbance measured between 190-260 nm. The spectra were subsequently averaged, and the averaged spectrum of the buffer was subtracted. Data acquired from CD were analysed using the online server DichroWeb (Whitmore and Wallace 2004, Whitmore and Wallace 2008).

2.4.6 Mass spectrometry

Mass Spectrometry (MS) can be used to identify protein compounds within a sample to elucidate their structural and chemical properties. The protein complexes were converted into gaseous ions, with or without fragmentation, and then characterised by their mass to charge ratios (m/z). Performing MS at near native conditions allows one to study proteins in their near-native states in solution, with partial oligomeric states retained.

The purified protein was buffer exchanged into 25 mM ammonium acetate using PD-10 desalting column (Amersham Biosciences) and diluted to 50 µM with the same buffer, and injected into a Liquid Chromatography MS (Agilent) located within the Analytical Laboratory of the Joseph Priestley building, Queen Mary University of London by nano-electrospray. Data were collected and analysed by RapidFire Integrator software (Agilent).

2.4.7 Chemical cross-linking

The cross-linking agent Dithiobis (succinimidylpropionate) (DSP) contains an amine-reactive N-hydroxysuccinimide (NHS) ester at each end of an 8-carbon spacer arm. These NHS esters react with primary amines, which can be found in the side chain of Lys residues and the N-terminus of each polypeptide, at pH 7 to 9, to form stable amide bonds, along with release of the NHS leaving group. In chemical cross-linking process, two or more protein monomers that are spatially next to each other can be 'linked' together, providing preliminary information of the quaternary association for transient semi-stable protein-protein interaction.

The purified protein was desalted into cross-linking reaction buffer (25 mM HEPES, 5 mM KCl, 5 mM $MgCl_2$ (pH 7.5)) using PD-10 desalting column (Amersham Biosciences). Chemical cross-linking was carried out in a total reaction volume of 20 µl in cross-linking reaction buffer with a variety of protein concentration. Cross-linking agent DSP (Thermo Scientific Pierce) was dissolved in Dimethyl Sulfoxide (DMSO, Thermo Scientific Pierce), and added to the reactions at a final concentration of 1 to 10 mM. Reactions were incubated at 23 ºC for 10 min and then stopped by addition of 50 mM Tris-HCl, 10 mM glycine (pH 7.5). Samples were loaded into a 10% non-reducing

separating gel (without reducing agent) for SDS-PAGE analysis, in order to retain disulfide bonds inside the cross-linking agent DSP and show the oligomeric state of Hik2.

2.4.8 Analytical ultracentrifugation

Analytical Ultracentrifugation (AUC) can be used to characterise various features of macromolecules in solution, including the size, shape, and interactions, by spinning and monitoring the samples in real time through an optical detection system. Two types of AUC experiments, the Sedimentation Velocity (SV) and Sedimentation Equilibrium (SE), and the data analysis were carried out by Ms. Liyana Azmi, under the supervision of Prof. Olwyn Byron at University of Glasgow.

Before AUC analysis, Hik2 protein was dialysed against the low salt buffer with glycerol (20 mM Tris-HCl, 10 mM NaCl, 5% glycerol, pH 7.5) and the high salt buffer with glycerol (20 mM Tris-HCl, 500 mM NaCl, 5% glycerol, pH 7.5). AUC was carried out in an Optima analytical ultracentrifuge (Beckman Coulter) at University of Glasgow. The SV data were collected at 49,000 rpm at 4 ºC. 360 µl of each sample, including 1.8, 1.3, 0.7, 0.3 mg/ml of Hik2 protein in low salt buffer, and 1.8, 1.0, 0.3 mg/ml of Hik2 protein in high salt buffer, were loaded into 3 mm path length charcoal-filled epon double-sector centrepieces. Data were acquired with absorbance optics and the scans were taken every 15 min for 25 h.

The SE experiments were performed at 4 ºC as well. 80 µl of a serial of Hik2 protein samples at concentrations of 1.8, 1.3, 0.7 mg/ml in both low salt and high salt buffers was tested, at the speed of 8,000 rpm and 12,000 rpm. Scans were taken every 3 h until

the equilibrium had been attained, indicated by the analysis of the scans by WinMATCH program (http://www.biotech.uconn.edu/auf/?i=aufftp).

US-SOMO (UltraScan Solution Modeler) was used to predict the different hydrodynamic parameters of Hik2 sample and provide atomic and lower-resolution bead model representations (Rai et al. 2005). SEDFIT and SEDPHAT software was used for the analysis of AUC data (Schuck 2000, Zhao et al. 2015).

2.5 Bioinformatics

Two bioinformatical methods were used in this thesis, in order to augment the homologous and structural information of CSK, Hik2 and the subdomain protein complexes, through computer analysis of their amino acid sequences.

2.5.1 Sequence alignment

The sequence alignment between known proteins is able to analyse amino acid sequences in order to identify similar regions, to acquire information of possible related structure, function, and evolutionary relationships. Homologous sequences were aligned with the Clustal Omega (Sievers et al. 2011) and displayed using the CLC sequence viewer 7 (QIAGEN).

2.5.2 Homology modelling

There is currently no crystal or NMR structure known for Hik2 or CSK, thus homology models of the full-length protein, or their subdomains, were generated using the modelling server Phyre2 (Protein Homology/analogY Recognition Engine) (Zhao et al.

2015) and SWISS-MODEL (Biasini et al. 2014), based on their amino acid sequences. I-TASSER (Zhang 2008) was also used to predict the structure of the sensor domain of CSK which was difficult to find homologous sequences. The one with the highest C-score (confidence score for estimating the quality of predicted models) among the generated five models was selected and shown in the following chapters.

2.6 Protein crystallisation studies

A few preliminary crystallisation screening trials have been performed with purified full-length *Synechocystis* Hik2, and the MBP- and SUMO-tagged sensor domains of Hik2 and CSK. Promising crystallisation conditions, for the full-length *Synechocystis* Hik2, were indeed identified and attempt to optimise these, for a larger scale production of this protein, were made.

2.6.1 Crystallisation screening

JCSG-plus, PACT premier and Morpheus Additive OptiMax kits (Molecular Dimensions) were used for the systematic screening of crystallisation conditions. 2-drop 96-well screening plates (Molecular Dimensions) were set up with a Mosquito robot (TTP labtech) controlled by the Mosquito software. Reservoir volumes were 80 µl and the volumes of two drops were 0.4 µl each, containing the reservoir and the protein solution with different concentrations at a volume ratio of 1: 1. The plates were incubated at 19 ºC in a temperature controlled room, or in a temperature controlled incubator (Molecular Dimensions), and inspected daily during the first week of filling the slab of each screen, and then every two weeks for up to three months.

2.6.2 Optimisation of crystallisation conditions

Optimisation of crystallisation conditions aids in the production of crystals of sufficient size and quality for diffraction studies, based on the information derived from the screening experiments. Once a crystal-like object was observed during the screening, the optimisation was carried out with 24-well XRL plates (Molecular Dimensions) set up for hanging drop. The chemical composition of the reservoir solution, the protein concentration, the volume ratio of the protein and the reservoir solution, and growth temperature, were included in the optimisation parameters.

2.7 Transmission electron microscopy

Transmission Electron Microscopy (TEM) is a biophysical technique enabling one to observe small objects, such as protein complexes. Together with computer-based single particle image averaging (analysis), the structural resolution obtained can ultimately be improved and extended in order to reconstruct three-dimensional (3D) structures through the back projection of two-dimensional (2D) images.

2.7.1 Specimen preparation

2.7.1.1 Protein concentration and dilution

Protein samples isolated previously were prepared for TEM through exchange in distilled water or buffer containing 20 mM Tris-HCl, 10 mM NaCl (pH 7.5), concentrated to a high concentration, typically > 1 mg/ml, and diluted into a series such as 5, 2.5 and 1.25 μg/ml. Further dilution may be required during the experiment depending on the observation at hand. The protein dilution method for the purpose of

TEM must be performed very gently so as to avoid inducing bubbles that will then interfere with the staining procedures below.

2.7.1.2 Negative staining

Negative staining is a commonly used method, which makes the specimen visible by staining the background, leaving the specimen untouched, and enhancing the edge-contrast of the protein complexes. It is technically simple to perform and provides some protection from electron beam damage, providing, as said, very high contrast for ease of direct specimen observation.

Purified and properly diluted protein samples were applied to the glow-discharged carbon-coated (thin-layer) copper 300-mesh EM grids (Agar, Ltd.) for 1 min. Then the excess sample was drained by a filter paper, from the reverse side of the grid, and the grid was dried in air for 1 min. Freshly prepared 2% Uranyl Acetate (UA, Agar Scientific), or Phosphotungstic Acid (PTA, Sigma) stain, was then applied in order to negatively stain the protein for a certain time length, typically ranging from 30 to 80 sec. Then the stain was drained, and the grid was exposed in air to dry for 3 min. Careful handling and repetition was required to ensure single particles were eventually spread evenly over the carbon support surface of the grid.

2.7.2 Image processing and single particle analysis

TEM sessions were performed within the NanoVision Centre at Queen Mary University of London. The images were recorded using the Charge Coupled Device (CCD) camera system (Olympus, Japan; model Morada 2k) attached to the 35-mm film port of a JEOL

1230 TEM, equipped with a tungsten filament and operating at 80,000 ×. This resulted in a sampling frequency of 5.962 Å per pixel at the specimen scale, however, a limitation of ~ 15 Å resolution was expected due to the presence of the UA or PTA negative stain (Kiselev et al. 1990). The Fourier-space power spectrum was calculated for each micrograph and, for the final analysis, many tens of micrographs were chosen from 5 sessions over 3 years for further single particle analysis, notably those that displayed minimal drift and astigmatism and having first minima at better than ~ 17 Å resolution. Single particle complexes were floated out into boxes of 64 × 64 pixels in size. Given that no correction for the Contrast Transfer Function (CTF) was applied, as the first minima were at the ~ 15 Å limitation expected given the use of negative stain, the final class averages presented were low band-pass filtered to ~ 20 Å resolution to be within a conservative range prior to interpretation.

Initial single particle images were selected using the 'boxer' module of EMAN2 (Tang et al. 2007), whose boxing algorithm was directed to automatically pick all possible single particles present, but not band-pass or normalise. The latter was to allow the Imagic-5 software environment (van Heel et al. 1996) to then be used for image normalisation, band-pass filtering, reference-free alignment and multi-variate statistical classification of the entire single particle image dataset accumulated.

Chapter 3 Structural characterisation and oligomerisation of full-length Hik2 protein and its subdomains

3.1 Overview

Structural characterisation methods are widely used to understand the function of a specific protein, and its protein-protein interaction mechanisms. Distinct functional and structural units in a protein, the subdomains, are usually studied independently. Since the kinase domain of *Synechocystis* Hik2 has been shown to be of low solubility (personal communication, Ibrahim I., 2016), other cyanobacterial strains were considered. *Thermosynechococcus elongatus* belongs to a thermophilic branch of cyanobacteria, and its genome has been sequenced (Nakamura *et al.*, 2002). Its proteins have been observed to be generally more thermodynamically stable during purification and easier to crystallise (Schubert *et al.*, 1997). Many protein crystal structures have been characterised from *Thermosynechococcus*, such as PS I, PS II, and the antioxidant effecting Dps protein (Jordan *et al.*, 2001; Schubert *et al.*, 1997; Zouni *et al.*, 2001; Franceschini *et al.*, 2006). In this fashion, *Thermosynechococcus elongatus* was chosen as a complementary cyanobacterial strain to *Synechocystis* sp. PCC 6803 in order to increase the likelihood of a positive outcome, such as gaining highly ordered crystals and maintaining particle homogeneity, for the structural studies performed below.

3.2 Results

In the following sections, the full-length *Synechocystis* Hik2 protein, the full-length *Thermosynechococcus* Hik2, the DHp domain and kinase domain of *Thermosynechococcus* Hik2 were expressed and purified, and size exclusion chromatography, dynamic light scattering, native-PAGE and cross-linking analysis

were performed in order to study whether or not the protein complexes were structurally stable and form higher order structures, e.g. oligomers, or were able to form self-forming lattices (2D or 3D), or perhaps prove to be labile, thus limiting the level of detail eventually possible given the quality of purification eventually obtained.

3.2.1 Production and purification of the full-length Hik2 protein and its kinase domain

3.2.1.1 Studies of the Hik2 from *Synechocystis* sp. PCC 6803

A colleague, Dr. Iskander Ibrahim, kindly provided training in molecular biology techniques, and had previously cloned the full-length *Synechocystis* sp. PCC 6803 Hik2 (GenBank: BAK49478.1), and transferred it to the BL21 strain of *E. coli* in the laboratory of Prof. John Allen. My work here has followed on from these initial studies by expressing and purifying the full-length Hik2 protein (Figure 3.1). From repeated studies, the Hik2 protein was observed to be present in the soluble fraction sample (Figure 3.1, lane 4), having an apparent molecular mass of approximately 50 kDa. This compared favourably with the calculated molecular mass of 48.6 kDa.

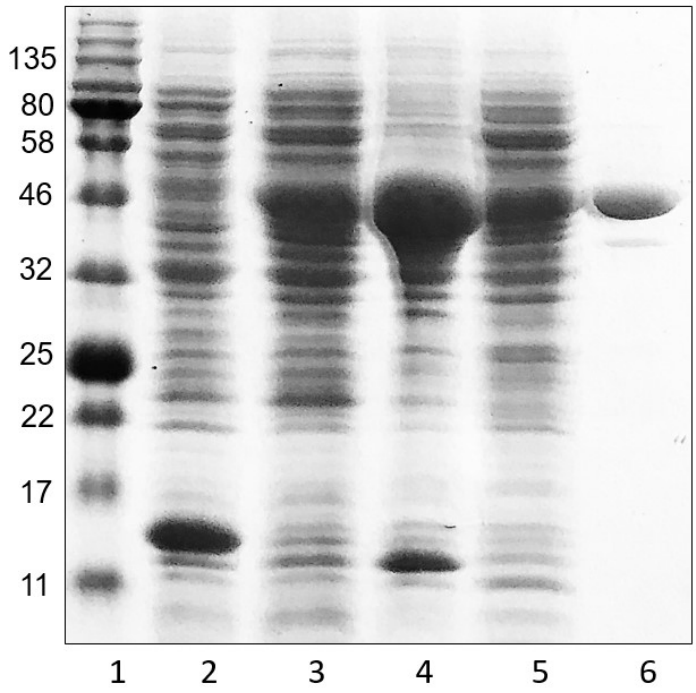


Figure 3.1 SDS-PAGE profile of *Synechocystis* Hik2 expressed and purified using Ni^{2+} affinity chromatography. Lane 1, protein molecular mass markers in kDa; lane 2, pre-induction sample of *E. coli* BL21 strain cell culture; lane 3, post-induction sample (BL21); lane 4, soluble fraction sample; lane 5, pellet sample; lane 6, elution sample of Hik2 protein from the Ni^{2+} column with buffer containing 20 mM Tris-HCl, 300 mM NaCl, 500 mM imidazole (pH 7.5). Samples were subjected to 12% SDS-PAGE followed by staining with Coomassie blue. The kDa values for the molecular mass markers are shown on the left.

In order to further study how domain interaction, within the core structure of the Hik2 protein complex, effects its function, the separate domains of Hik2 were expressed and purified sequentially, however, the DHp domain (necessary for structural dimerisation) and the truncated form (kinase domain, for kinase specific structural studies) of *Synechocystis* Hik2 unfortunately proved to be stubbornly insoluble (data not shown), and studies involving these latter two domains were not pursued.

3.2.1.2 Studies of Hik2 from *Thermosynechococcus elongatus*

A different species of cyanobacteria, *Thermosynechococcus elongatus*, obtained by Dr. Ibrahim as a cell culture, as a kind gift from Dr. Dennis J. Nuernberg (sourced from a hot spring in Beppu, Japan), was used to clone these two subdomains of Hik2, the DHp and the truncated domain (see section 2.1). These two specific subdomain architectures from *Thermosynechococcus* Hik2 are shown in Figure 3.2 A. Furthermore, as an interesting potential future evolutionary comparison, it may be noted that the truncated form of CSK, from the green plant *Arabidopsis thaliana*, was also cloned by Dr. Iskander Ibrahim shortly before these Hik2 studies started. The overview to that work is shown in Figure 3.2 B.

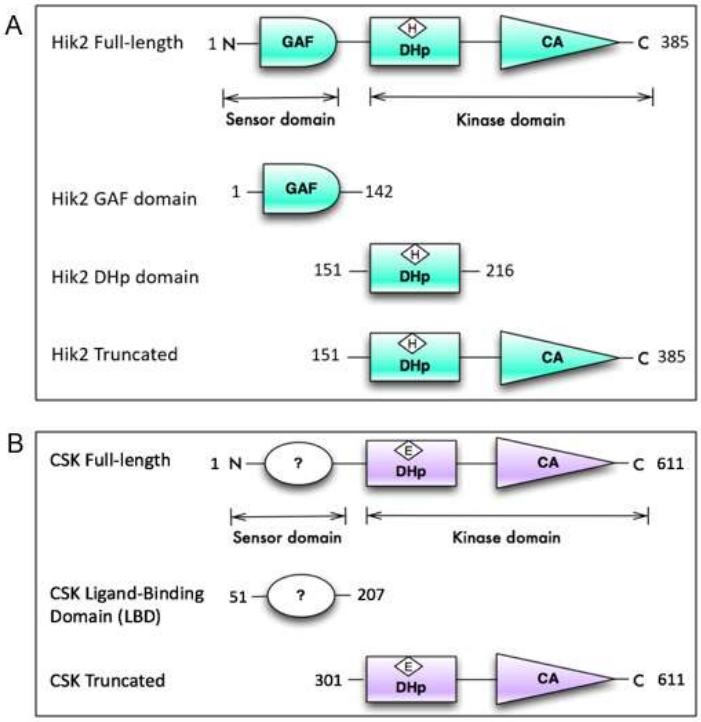


Figure 3.2 Domain architecture of *Thermosynechococcus* Hik2 and *Arabidopsis* CSK. A) Domain architecture of the full-length and subdomains of *Thermosynechococcus* Hik2. B) Domain architecture of the full-length and subdomains of the *Arabidopsis* CSK. Numbers of the starting and ending amino acids are marked, as are the key sensor and kinase domain locations.

The full-length *Thermosynechococcus hik2* gene (1158 base pairs (bp)) was cloned and ligated using a pET_21b vector, and test digested (Figure 3.3). The restriction enzyme NdeI, and the isocaudomers SalI and XhoI were used in the digestion. NdeI and SalI were used to digest the *hik2* gene PCR product, while NdeI and XhoI, the pET_21b vector. Two identical overhanging termini sequences (TCGAC and TCGAG) were generated by the isocaudomers SalI and XhoI respectively, and then ligated to one another, but could not be cleaved by SalI or XhoI again. The reason that the arrowed band in Figure 3.3 appears smaller than the expected 1158 bp is due to the presence of an XhoI restriction site at the 802 bp position of the *hik2* DNA sequence which was cleaved during the test digestion. Meanwhile, the resulting sequence ligated by the SalI-cut and XhoI-cut site could not be recognised and cleaved by XhoI.

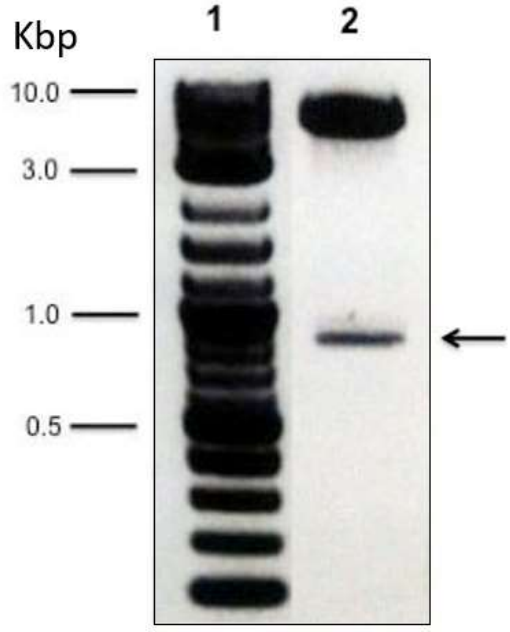


Figure 3.3 Digestion of the *Thermosynechococcus hik2* ligated with pET_21b vector. Lane 1, 2-log DNA ladder marker; lane 2, digested *hik2* full length DNA with pET_21b vector. The molecular mass markers are shown on the left in Kilo base pairs (Kbp). Samples were subjected to separation with 1% agarose gel. The plasmid was digested by NdeI and XhoI. The arrow indicates the XhoI digested *hik2* gene band at the 802 bp position.

Following on, the construct was used as a template for PCR amplification of the GAF domain (sensor domain), DHp (Dimerisation and phosphor acceptor) domain, CA (Catalytic and ATP binding) domain and the truncated form of *Thermosynechococcus hik2* gene (Figure 3.4).

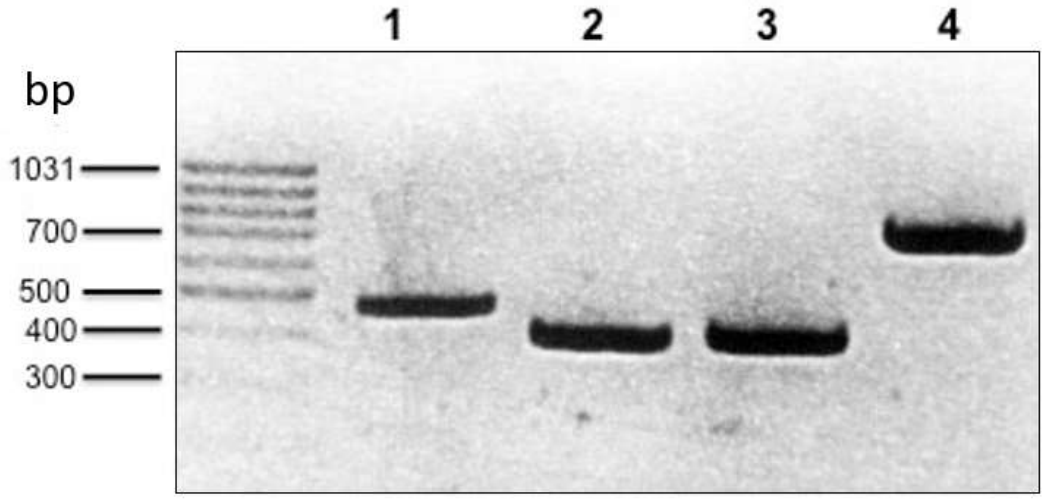


Figure 3.4 PCR amplification of the GAF, DHp, CA domain and the truncated form of *Thermosynechococcus hik2*. Lane 1, the PCR product, in sequence, of the GAF domain (447 bp in length); lane 2, the DHp domain (384 bp); lane 3, the CA domain (375 bp); lane 4, the truncated form (732 bp). Samples were subjected to a separating gel of 1% agarose. The molecular mass markers, in the form of base pairs, are shown on the left.

The full-length *Thermosynechococcus* Hik2 and its subdomain fragments were ligated with pET_21b vector, overexpressed in *E. coli* and purified using a Ni^{2+} column. Figure 3.5, lane 8 shows the purified Thermo_Hik2F; lane 9, Thermo_Hik2T; lane 10 Thermo_DHp. The apparent molecular masses were found to be: Thermo_Hik2F, 45 kDa; Thermo_Hik2T, 29 kDa; Thermo_DHp, 16 kDa; all corresponding to their calculated molecular mass, as confirmed by ExPASy - Compute pI/Mw tool (http://web.expasy.org/compute_pi/) (Gasteiger et al. 2003).

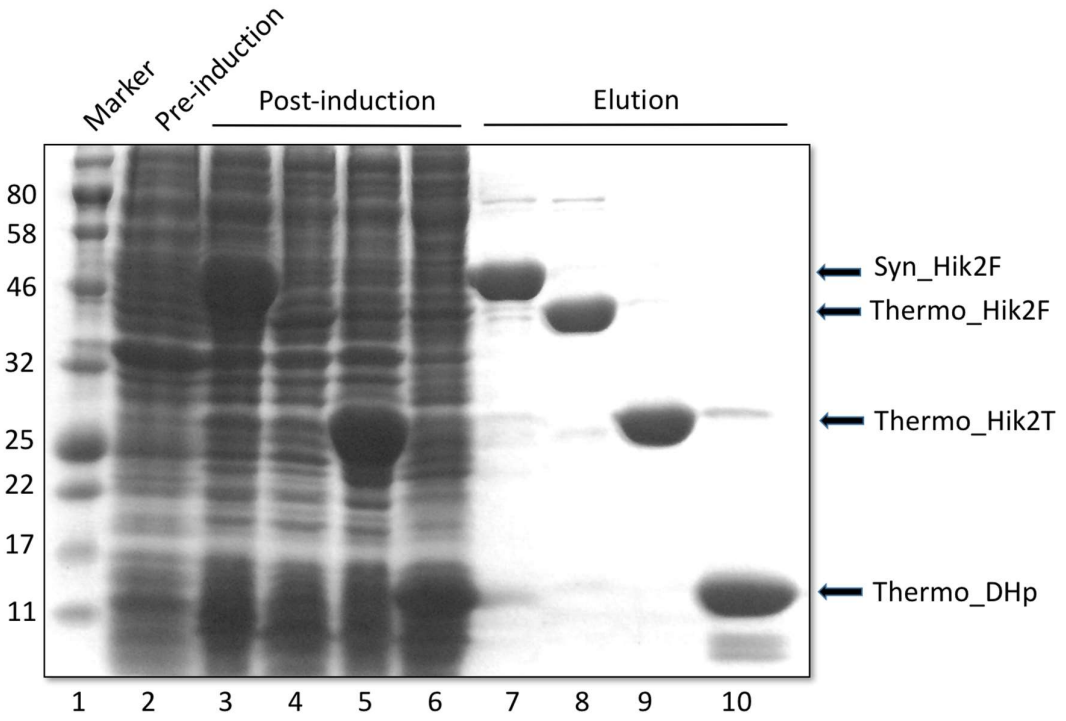


Figure 3.5 Characteristic SDS-PAGE profile of Hik2 and its subdomain proteins, with respect to over-expression and purification. Lane 1, protein molecular mass markers; lane 2, pre-induction sample of *E. coli* cell culture; lane 3 to 6 are total cell lysate for post-induction: lane 3, full-length *Synechocystis* Hik2 (Syn_Hik2F); lane 4, *Thermosynechococcus* full-length Hik2 (Thermo_Hik2F); lane 5, *Thermosynechococcus* truncated form (Thermo_Hik2T); lane 6, *Thermosynechococcus* DHp domain (Thermo_DHp); lane 7 to 10 correspond to elution fractions: lane 7, Syn_Hik2F; lane 8, Thermo_Hik2F; lane 9, Thermo_Hik2T; lane 10, Thermo_DHp. Samples were subjected to a 12% separating gel for SDS-PAGE followed by staining with Coomassie blue. The positions of the overexpressed proteins are indicated by horizontal arrows on the right.

The GAF domain and CA domain of *Thermosynechococcus* Hik2 were not soluble (data not shown), since the surface of these two domains may be hydrophobic. In Chapter 4, attempts were made to solubilise the GAF domain.

3.2.2 The full-length and subdomains of the Hik2 protein exist in different oligomeric states under conditions of low salt

The *Synechocystis* full-length Hik2 sample in low salt conditions (25 mM ammonium acetate) was probed using nano-electrospray ionisation Mass Spectrometry (MS, Figure 3.6) within the Analytical Laboratory of Queen Mary, University of London (see section 2.4.6). Mass determinations for the full-length Hik2 protein showed the molecular mass of a monomer to be 48,533 Da, consistent with a calculated molecular mass of 48,640 Da for Hik2 protein and a his-tag. As shown in Figure 3.6, different oligomeric states were detected under these conditions.

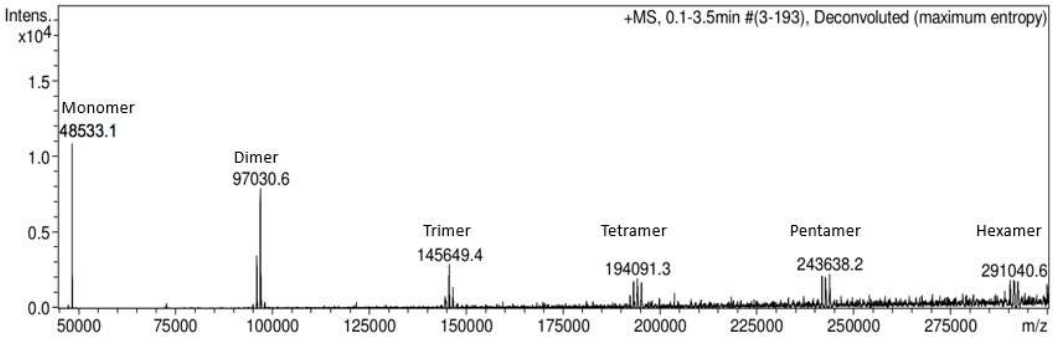


Figure 3.6 Mass spectrometry analysis of *Synechocystis* full-length Hik2 protein. Sample analysed by nano-electrospray ionisation in an aqueous solution containing 25 mM ammonium acetate. Oligomeric states observed were as labelled, the monomer (48,533.1 Da), dimer (97,030.6 Da), trimer (145,649.4 Da), tetramer (194,091.3 Da), pentamer (243,638.2 Da), hexamer (291,040.6 Da).

It was suggested that native-MS could provide solution phase structural information and as such for protein noncovalent interactions (Benesch and Ruotolo 2011). However, the presence of trimer and pentamer was initially anticipated to be unlikely because Hik2 was expected to dimerise due to the dimerisation domain. It should be noticed that the nano-electrospray ionisation MS is not a native-MS, and the condition required for

particles to ionise are hostile for many protein complexes, which may be the reason that Hik2 oligomers were being disordered during the electrospray ionisation step. Thus, non-gas phase methods, such as gel filtration or native-PAGE were considered for the initial characterisation of the protein complexes.

Size exclusion chromatography was applied to further investigate the oligomeric state of Hik2 and its subdomains. Protein was loaded onto a Superdex 200 column that calibrated and ran with a low salt buffer (20 mM Tris-HCl, 10 mM NaCl (pH 7.5)), where both the full-length Hik2 protein from *Synechocystis* (48.5 kDa) and *Thermosynechococcus* (45 kDa), eluted as a possible hexameric form according to their apparent molecular mass (Figure 3.7). It was also observed that the truncated form of *Thermosynechococcus* Hik2 (28 kDa) eluted as a possible hexamer. Moreover, the DHp domain (15 kDa) forms a higher order oligomer on its own (Figure 3.7, panel B).

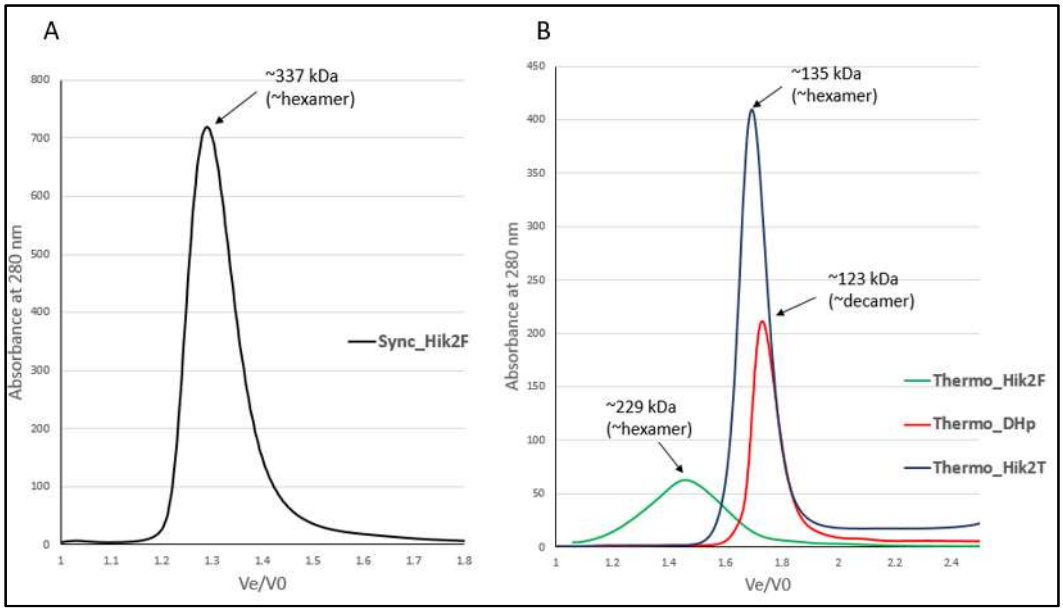
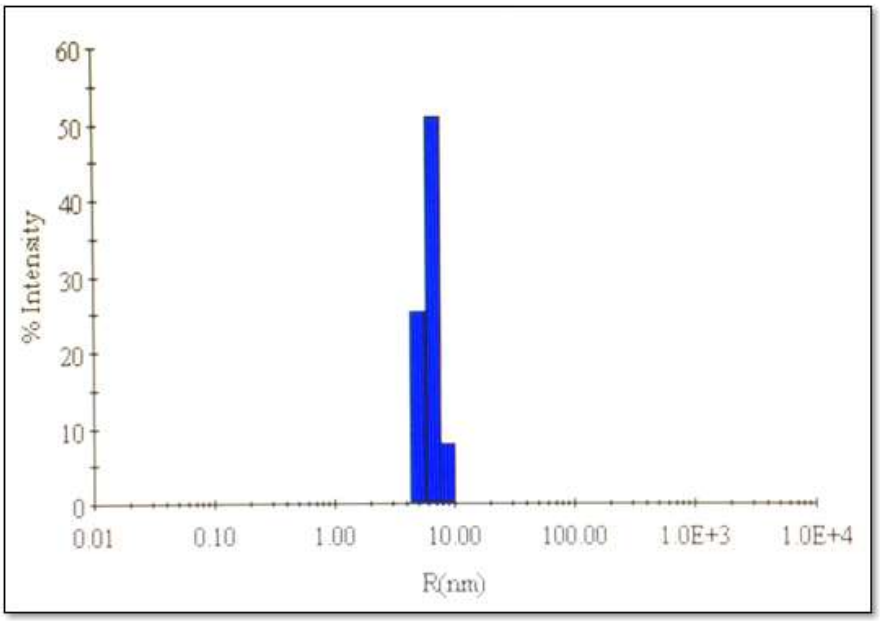


Figure 3.7 A typical elution profile from a Superdex 200 column eluted with low salt buffer. A) *Synechocystis* full-length Hik2 protein profile. B) *Thermosynechococcus* full-length Hik2, its DHp domain and kinase domain protein profiles.

Dynamic Light Scattering (DLS) was used to analyse the purified samples in order to reveal their molecular mass in solution. The observed result (Figure 3.8) conformed that the *Synechocystis* full-length Hik2 (48.5 kDa) formed a hexamer (~ 263 kDa) under the low salt conditions of 20 mM Tris-HCl, 10 mM NaCl (pH 7.5).

A



B

	R (nm)	%Pd	MW-R (kDa)
Syn_Hik2F	6.438	16.8	263

Figure 3.8 DLS profiles of *Synechocystis* full-length Hik2 in low salt buffer. A) The intensity distribution of the particle size. B) The corresponding polydispersity (Stepto 2010) and molecular mass of the particle.

Clear native-PAGE (Figure 3.9 A) was used to confirm the oligomeric state of the proteins, with the support of Dr. Ibrahim. Here, it was observed that the full-length Hik2, from both species, and the *Thermosynechococcus* truncated Hik2, separated into several bands on native gels, no matter whether it was eluted by Ni^{2+} affinity chromatography or from the peak elution via size exclusion chromatography, while the DHp domain ran as only one band.

Since the migration distance of the protein depends on the protein charge, protein size and the pore size of the clear native gel, gradient native-PAGE was also applied (Figure 3.9 B). In a gradient native gel, the mobility of a protein decreases as migrating through the increasing acrylamide concentration, into regions with smaller pore sizes. Eventually, each protein will reach a minimum migration rate, which is constant at its pore limit, and the relative position reflects the molecular mass directly. Interestingly, a single band can be observed on the gradient native gel, for full-length Hik2, truncated Hik2 protein, and the DHp domain, which means the majority of proteins in the samples have the same molecular mass. The separation of protein in the clear native gel in Figure 3.9 A may be caused by the presence of a different mass/charge ratio, due to proteins with different conformations.

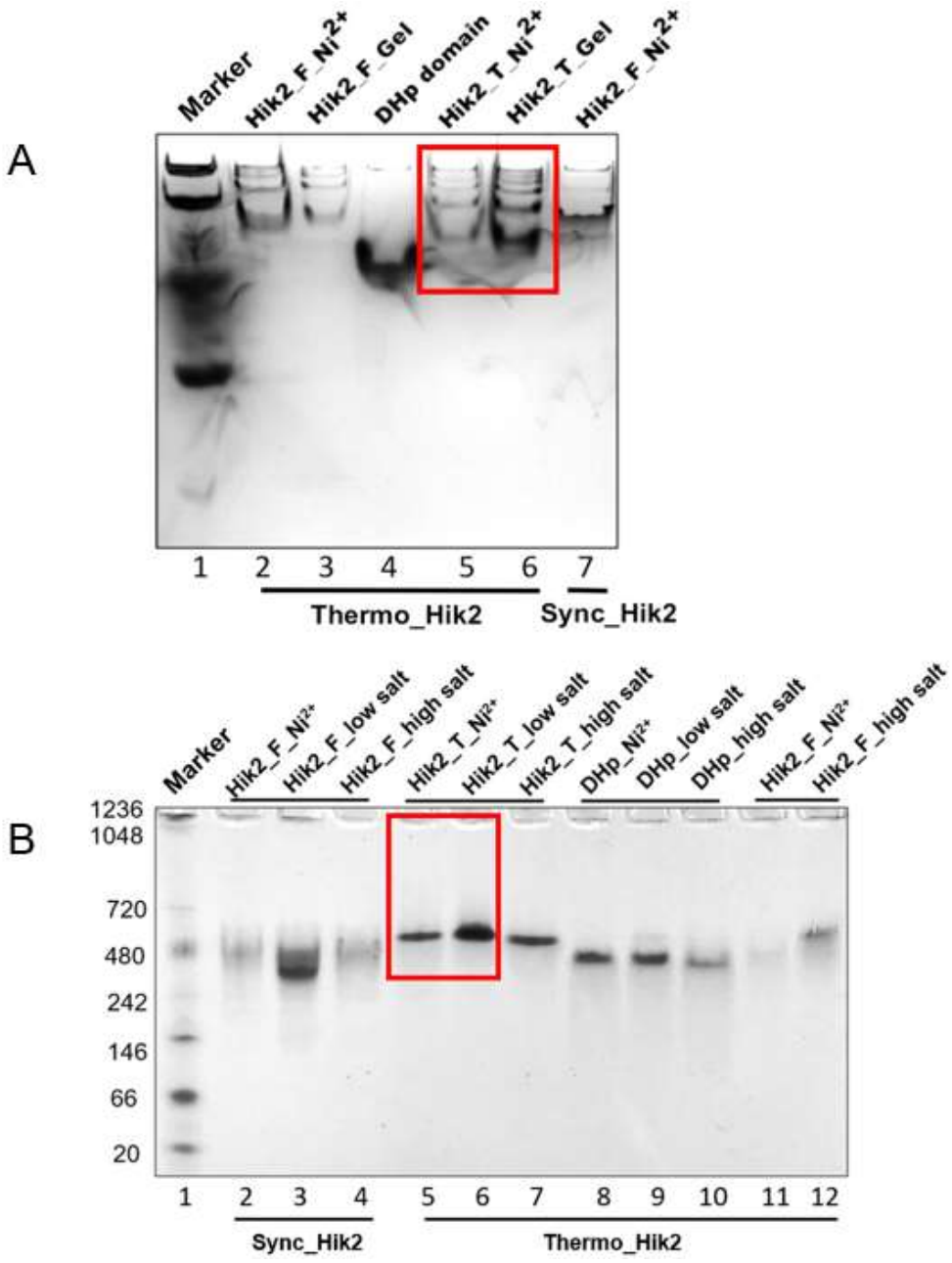


Figure 3.9 Native-PAGE profiles of Hik2 and its subdomains. A) Clear native-PAGE profile. Lane 1, native-PAGE protein markers; lanes 2 to 6, as labelled, samples from *Thermosynechococcus*; lane 7, as labelled, Hik2 sample from *Synechocystis*. B) 4% to 20% gradient native-PAGE profile. Lane 1, native-PAGE protein molecular mass markers in kDa; lanes 2 to 4, as labelled, Hik2 sample from *Synechocystis*; lanes 5 to 12, as labelled, samples from *Thermosynechococcus*. The same samples of Thermo_Hik2T from both Ni^{2+} column elution and gel filtration elution are highlighted in red boxes in the two native-PAGE profiles.

Chemical cross-linking was performed together with Dr. Ibrahim to study the interactions between protein molecules. In order to determine whether the oligomeric state of Hik2 is dependent on the concentration of the cross-linker Dithiobis (succinimidylpropionate) (DSP), Hik2 protein samples were incubated with a variety of concentrations of DSP, for 10 min at 23 ºC. Cross-linked products were then resolved by non-reducing SDS-PAGE. Figure 3.10 depicts the gel profile, and notably lane 2 revealed the untreated Hik2 protein to migrate with an apparent molecular mass of ~ 50 kDa, thereby corresponding to the monomeric form of Hik2 protein. Lanes 3 to 10 reveal chemical cross-linking products for four distinct protein bands. These correspond to monomers, as per lane 2, and a second band just below 190 kDa, is most likely to indicate a tetrameric form. Two further bands above 250 kDa provide for the possibility of even higher-oligomers, possibly hexameric and octameric complexes. Although the oligomerisation state of Hik2 was not affected by increasing the concentration of DSP to 3 mM (Figure 3.10, lanes 2 to 5), augmenting the concentration to be above 3 mM resulted in a decrease in both monomeric and higher-order oligomers (Figure 3.10, lanes 6 to 12). Given this, from hereon only 2 mM of DSP was used in subsequent experiments.

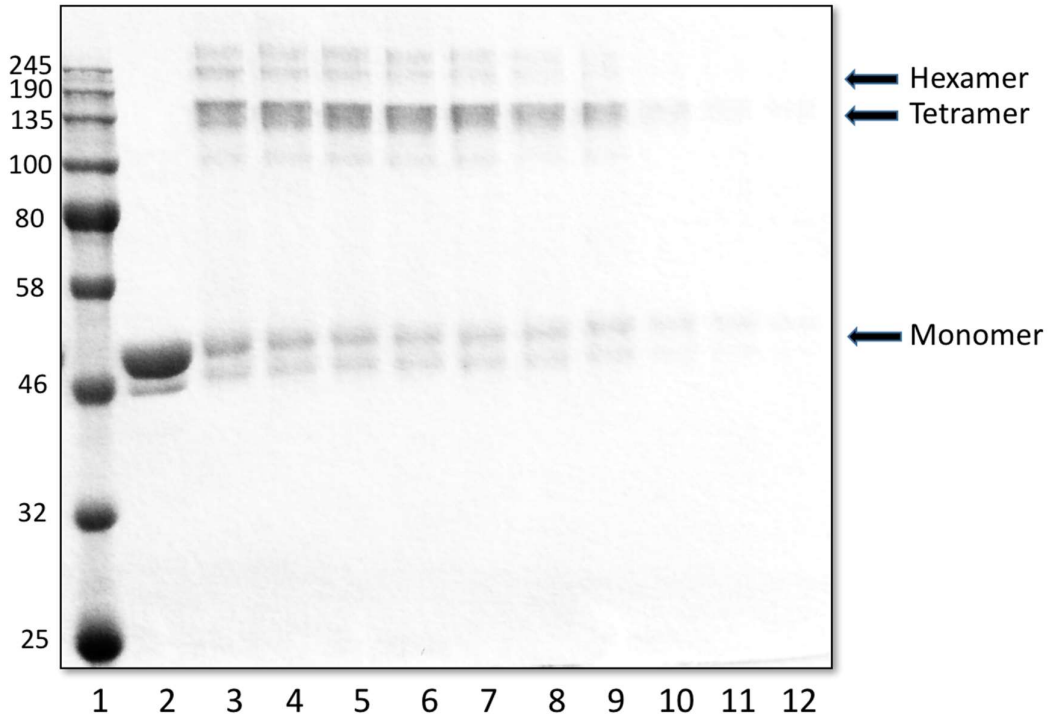


Figure 3.10 SDS-PAGE profile of chemical cross-linking of *Synechocystis* full-length Hik2. Lane 1, protein molecular mass markers; lane 2, untreated Hik2 protein (control); lanes 3 to 12, Hik2 treated with 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 mM DSP, respectively. Samples were subjected to a 10% non-reducing separating gel during SDS-PAGE. Molecular mass markers are shown on the left in kDa. The possible oligomeric states of Hik2 are indicated as per the arrows on the right.

3.2.3 NaCl affects the oligomeric state

Size exclusion chromatography revealed that when adding 500 mM NaCl to the buffer, the Hik2 full-length and truncated proteins converted from an apparent molecular mass of 337 kDa, an expected hexamer, to tetramer (204 kDa), and interestingly the DHp domain converts from an apparent decamer to being hexameric form (see and compare Figure 3.11 with Figure 3.7).

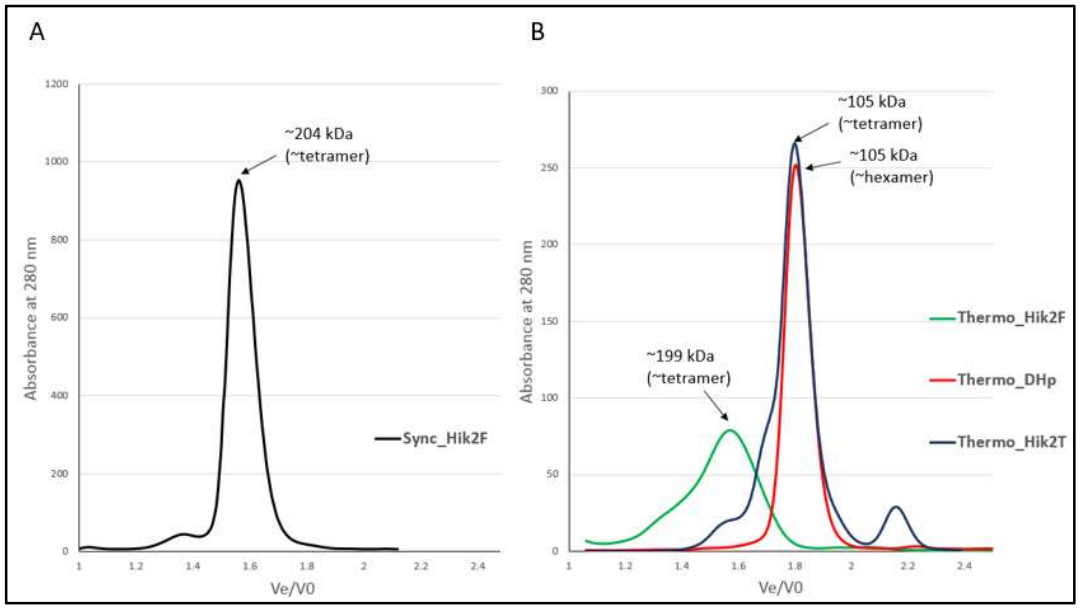
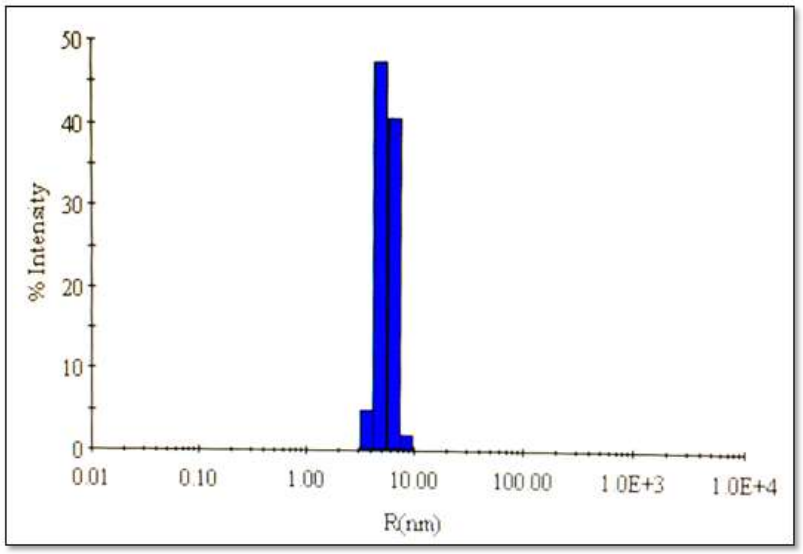


Figure 3.11 A characteristic elution profile from a Superdex 200 column with high salt buffer. A) *Synechocystis* full-length Hik2 protein profile. B) *Thermosynechococcus* full-length Hik2, the DHp domain and the truncated Hik2 protein profiles.

DLS further demonstrated that the introduction of 500 mM NaCl converts the full-length *Synechocystis* Hik2 protein from being a hexamer to tetrameric in molecular mass (see and compare Figure 3.12 with Figure 3.8).

A



B

	R (nm)	%Pd	MW-R (kDa)
Syn_Hik2F	5.557	17.4	186

Figure 3.12 DLS profiles for *Synechocystis* full-length Hik2 in high salt buffer. A) The intensity distribution of the particle size. B) The corresponding polydispersity (Stepto 2010) and molecular mass of the particle.

When Hik2 elution samples, from Superdex 200 column size exclusion chromatography, using both low salt and high salt buffers, were applied to Vivaspin 6 protein concentrators having a MWCO (molecular weight cut-off) of 50 kDa, it was found that the flowthrough of protein sample, for the low salt buffer, did not contain traceable Hik2 protein, while it was detectable in small amounts in the flowthrough of protein in high salt buffer. This suggests some lower order Hik2 oligomers were escaping past the filter.

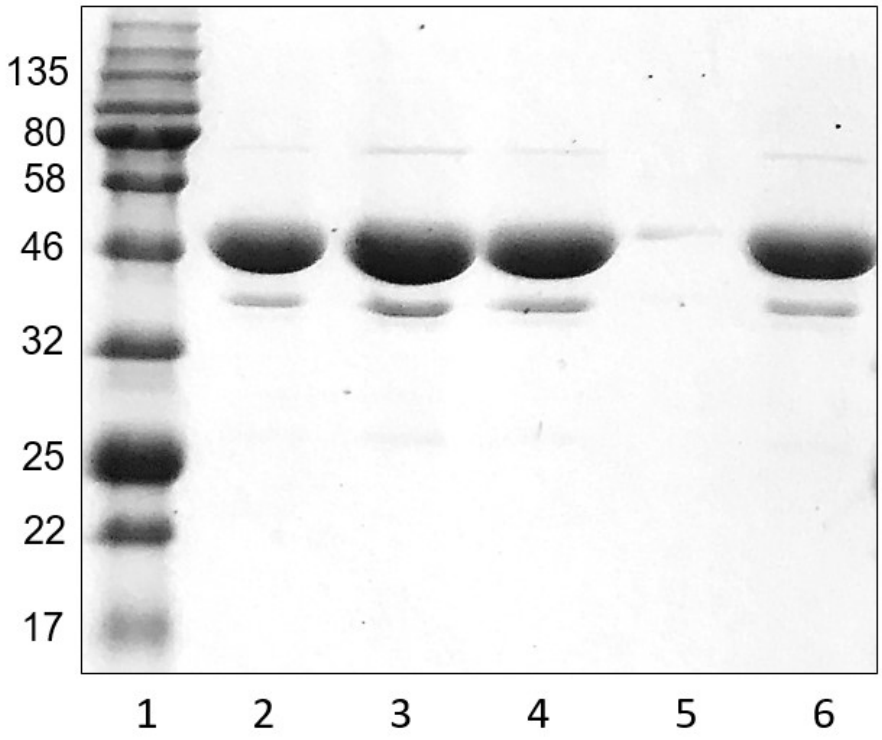


Figure 3.13 SDS-PAGE profile of *Synechocystis* full-length Hik2 protein samples after concentration. Lane 1, protein molecular mass markers in kDa; lane 2, Hik2 elution sample from Ni^{2+} column. For lanes 3 to 6, the Hik2 samples were subjected to Vivaspin 6 protein concentrator (MWCO 50 kDa), either in lane 3 the concentrated sample in low salt buffer; lane 4, the concentrated sample in high salt buffer; lane 5, the flow-through sample in low salt buffer; lane 6, the flow-through sample in high salt buffer. Samples were subjected to SDS-PAGE with a 12% separating gel followed by staining with Coomassie blue.

The result obtained from chemical cross-linking also indicated that addition of 500 mM NaCl increased the amount of the tetrameric state and reduced the amount of hexamer present (Figure 3.14).

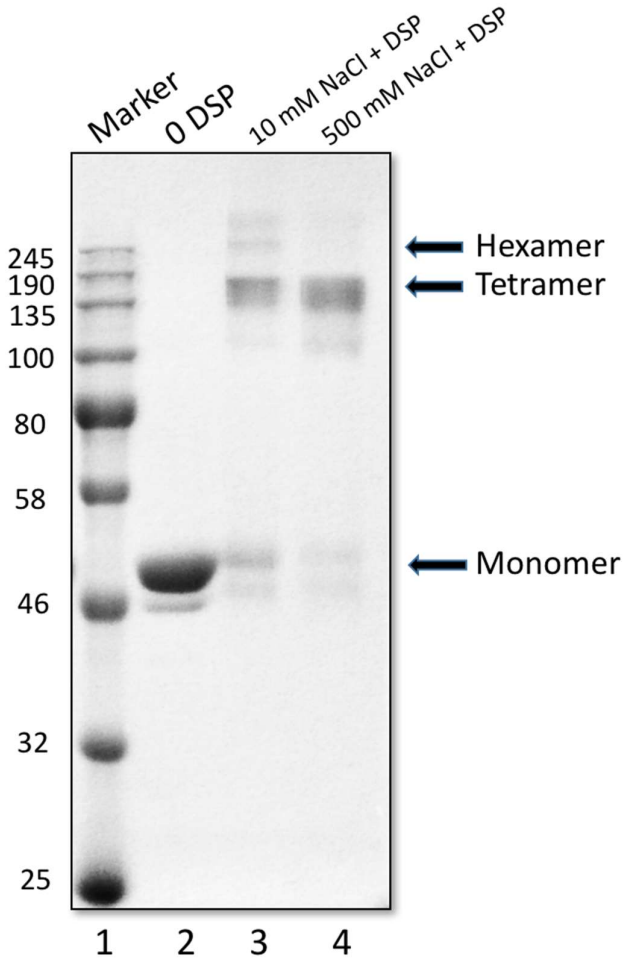


Figure 3.14 SDS-PAGE profile showing effects of chemical cross-linking with NaCl on the oligomeric state of *Synechocystis* full-length Hik2. Lane 1, protein molecular mass markers in kDa; lane 2, untreated Hik2 protein (control); lane 3, Hik2 protein in buffer containing 20 mM Tris-HCl, 10 mM NaCl (pH 7.5) treated with 2 mM DSP; lane 4, Hik2 protein in buffer containing 20 mM Tris-HCl, 500 mM NaCl (pH 7.5) treated with 2 mM DSP. Samples were subjected to a non-reducing separating gel of 10% acrylamide. The possible oligomeric states of Hik2 are indicated on the right.

3.2.4 Circular dichroism spectroscopy analysis of *Synechocystis* full-length Hik2 protein

3.2.4.1 The principle of circular dichroism spectroscopy

The secondary structure content of a protein gives important characteristic information regarding its structure. The proportions of α-helices, β-sheets and random coils form the basis for most classifications of protein folds. Circular Dichroism (CD) spectroscopy is a type of light absorption spectroscopy that can be used to gather information about this protein secondary structure and its folding properties. It can measure the difference in absorbance of left-handed and right-handed circularly polarised light, instead of the isotropic light, by a protein. The 'far-UV' spectral region (190 to 250 nm) is usually used to determine the secondary structure by CD spectroscopy, because the chromophore is the peptide bond between these wavelengths, and characteristic CD spectrums are known for α-helix, β-sheet, and random coil structures respectively (Figure 3.15).

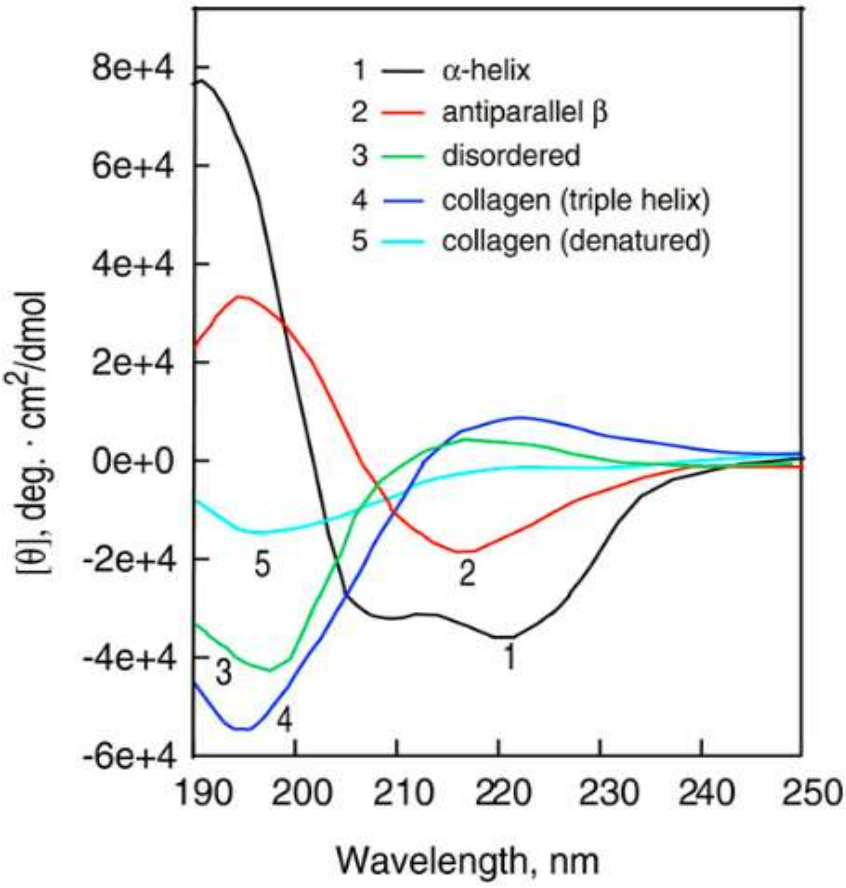


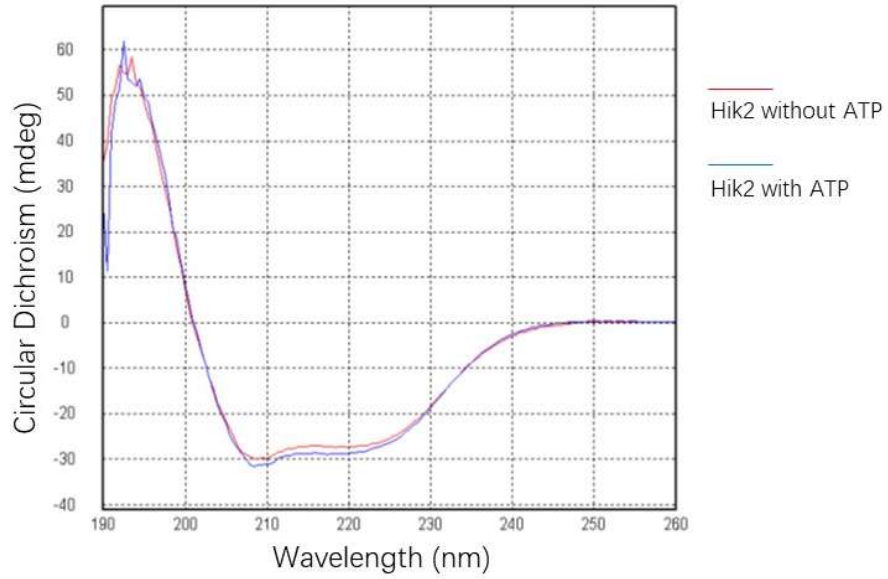
Figure 3.15 Circular dichroism spectra of polypeptides and proteins with representative secondary structures. The poly-L-lysine shows α-helical conformation at pH 11.1 (black), antiparallel β-sheet at pH 5.7 (red), and random coil at neutral pH (green). The blue and cyan curves represent the placental collagen in its native triple-helical and denatured forms. Figure adapted from (Greenfield 2006).

CD spectroscopy is a non-destructive experimental method, which can be used to study proteins in solution close to the native conditions, and only a small amount of protein is needed. However, there are also limitations, such as influences from aromatic sidechains, or prosthetic groups, which make the secondary structure analysis less accurate.

3.2.4.2 The secondary structure information of the full-length *Synechocystis* Hik2 protein as measured by CD spectroscopy

CD spectroscopy was used to collect the secondary structure information of the full-length *Synechocystis* Hik2 protein. Since Hik2 has autophosphorylation activity, ATP was then added to both the buffer and the protein sample to test if the secondary structure changes after Hik2 binds to ATP and being phosphorylated. The data collected were analysed by the online server DichroWeb (Whitmore and Wallace 2004, Whitmore and Wallace 2008), and the results are shown in Figure 3.16. It can be seen that Hik2 is a folded protein with a large number of α-helices, and also a great amount of random coil.

A



B

	Alpha helix	Beta sheet	Random coil
Hik2 without ATP	0.29	0.17	0.54
Hik2 with ATP	0.32	0.18	0.50

Figure 3.16 CD data analysed by DichroWeb for the full-length *Synechocystis* Hik2. A) The CD spectrum. B) The estimate of the relative content of secondary structure elements. The NRMSD of the calculation for the Hik2 without ATP is 0.056, and the Hik2 with ATP, 0.054.

It showed similar CD spectra and the relative content of the secondary structure elements with or without ATP in the samples, indicating that the secondary structure does not change much. This is predictable because studies on other histidine kinases have proven that the ligand binding and autophosphorylation usually relate to conformational changes, which does not affect the secondary structure (Casino et al. 2010).

3.2.5 Analytical ultracentrifugation analysis on *Synechocystis* full-length Hik2 protein

3.2.5.1 The principle of analytical ultracentrifugation

Analytical Ultracentrifugation (AUC) can be broadly applied to characterise macromolecules (e.g. proteins, viruses) in solution, such as the size and shape of molecules, as well as the interactions. It does not require the samples to be labelled or modified by other chemicals, or interacted with any matrix or surface, but simply uses a centrifugal force and the real-time observation of the resulting spatial redistribution of the investigated molecules. Three optical systems can be used for analytical ultracentrifugation, absorbance, interference, and fluorescence, and interchanged depending on the different molecules to be observed. The samples are usually run in a vacuum environment, which helps to reduce the friction generated at high speeds and maintain a constant temperature.

There are two principal types of experiments carried out in AUC, the Sedimentation Velocity (SV) and Sedimentation Equilibrium (SE), as exemplified by a protein of 50 kDa with a sedimentation coefficient of 4 S in Figure 3.17 (Cole et al. 2008). The SV

experiments are carried out in a cell with two channels, loaded with the protein solution in one, and the reference buffer in another (Figure 3.17 A). A series of sample concentrations are often applied for SV experiments, and a high rotor speed is required. During the SV run, scans are recorded at fixed intervals, such as the sample scanned and displayed at 20 min intervals by different colours, with a rotor speed of 50,000 rpm, as in Figure 3.17 A. The centrifugal force forces the protein to move towards the bottom of the cell, producing a depletion of solute near the meniscus and the formation of a sharp boundary between the depleted region and the uniform concentration. Two main parameters, the sedimentation coefficient (s) and the translational diffusion coefficient (D) can be calculated by analysing the boundary movement over time, under the consideration of the gravitational force, the frictional force and the diffusion of the molecules opposing the concentration gradient (Campbell 2012).

The sedimentation coefficient (s) is an important parameter, with a unit of Svedberg (S) (1 S = 10^{-13} seconds) that can provide information on the size and shape of the molecule. It is defined to be the ratio of a particle's sedimentation velocity to the acceleration that is applied to it. Thus it tells how fast a protein moves in solution, and it can be inferred that an increasing sedimentation will appear linked to an increase in protein mass, while the opposite is the case; increasing the size or asymmetry of a protein will decrease its sedimentation (Erickson 2009).

When performing an SE experiment (Figure 3.17 B), the centrifugal force should be sufficiently small, to allow an equilibrium concentration of the macromolecules to distribute throughout the cell. At this time, the flux due to sedimentation is exactly

balanced by the flux due to diffusion (Campbell 2012). A more accurate molecular mass can be calculated, and the oligomeric state is possible to infer consequently.

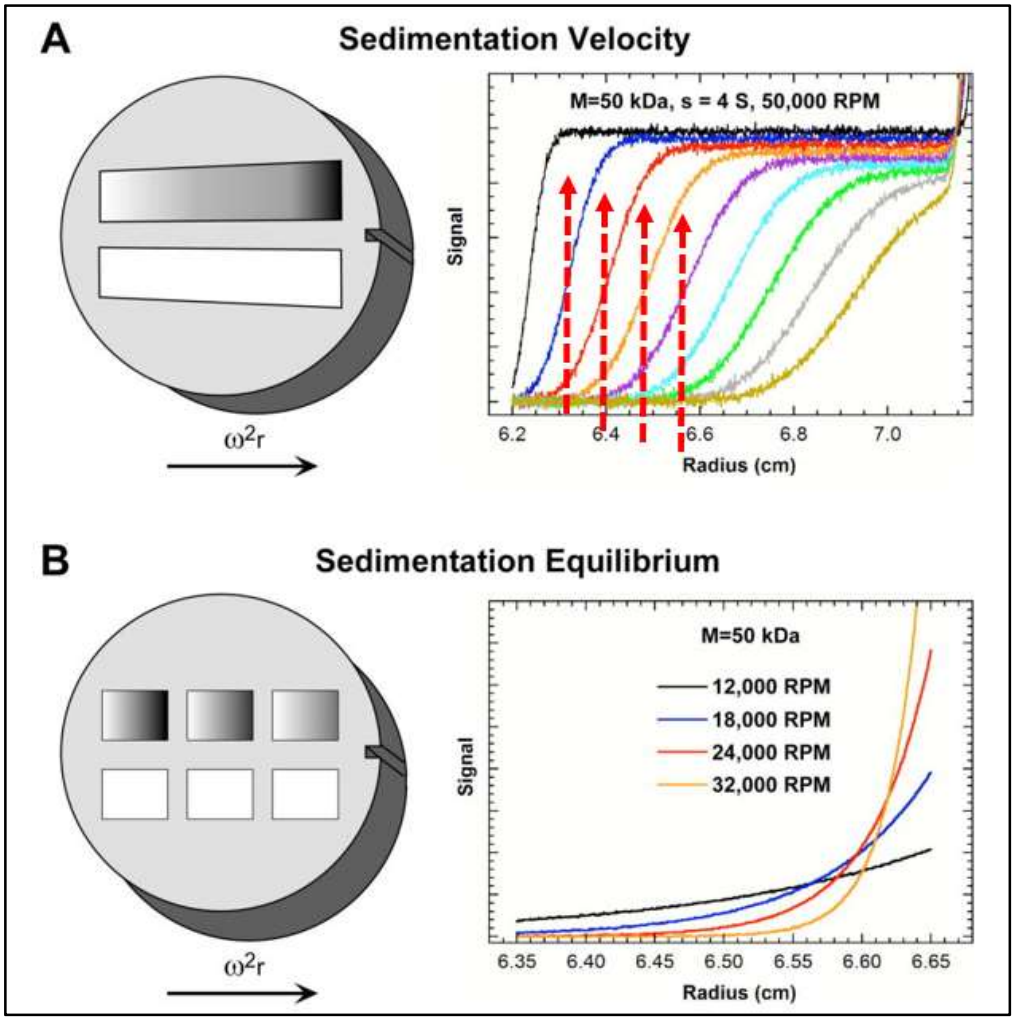


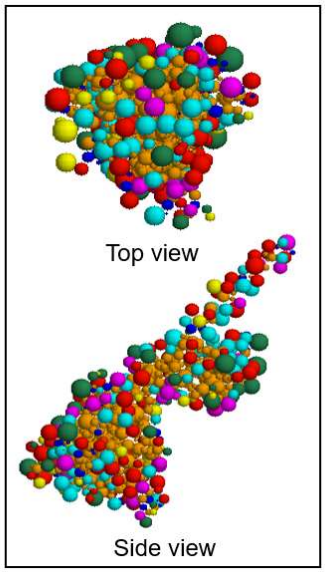
Figure 3.17 An example of basic analytical ultracentrifugation experiments. A) The SV experiment performed in a cell with two channels, and scans were recorded at 20 min intervals with a rotor speed of 50,000 rpm. Red dashed arrows indicate the midpoint positions of the boundaries. B) The SE experiments. Equilibrium measurements usually employ a six-sector cell with three loading concentrations. The equilibrium concentration gradients were simulated for four rotor speeds ranging between 12,000 and 32,000 rpm. Figure adapted from (Cole et al. 2008).

3.2.5.2 Prediction of hydrodynamic parameters for *Synechocystis* Hik2

US-SOMO (UltraScan Solution Modeler) is a software that can process atomic and lower-resolution bead model representations of a protein or other macromolecules and predict different hydrodynamic parameters including sedimentation and diffusion coefficients (Rai et al. 2005). Before the AUC experiment, the I-TASSER predicted structure of Hik2 protein was applied to US-SOMO to calculate the bead model and the hydrodynamic parameters.

As is seen in Figure 3.18 A, the side view of the bead model for the predicted *Synechocystis* Hik2 monomer shows a quite elongated structure. A molecule with an elongated shape will experience more friction from the solvent in AUC, so it may tend to sediment slower, resulting in a smaller calculated molecular mass than its true molecular mass. The hydrodynamic parameters of the Hik2 monomer were calculated by US-SOMO modeler (Rai et al. 2005), and shown in Figure 3.18 B. The predicted $s_{20,w}$ value (the sedimentation coefficient of a particle in water at 20 ºC) is 3.34 S, which was then used to compare to the actual experimental data later.

A



B

SOMO Hydrodynamic Results (Water at 20°C) (Density 1.00194 cP, Viscosity 0.998234 g/ml)	
Model:	Hik2_model_ITASSER-A10R30syOThyG5-a2b
Method:	SMI
Total Beads in Model:	661
Used Beads in Model:	551
Molecular Mass:	4.7561e+04 Da
Part. Specif. Volume:	0.738 cm^3/g
Sedimentation Coefficient s:	3.34e+00 S
Tr. Diffusion Coefficient D:	6.49e-07 cm/sec^2
Stokes Radius:	3.30e+00 nm
Frictional Ratio:	1.37
Radius of Gyration:	3.21e+00 nm
Relaxation Time, tau(h):	4.49e+01 ns
Intrinsic Viscosity:	5.50e+00 cm^3/g

Figure 3.18 The predicted bead model and hydrodynamic parameters of *Synechocystis* Hik2. The US-SOMO software and the *Synechocystis* Hik2 PDB file predicted by I-TASSER were used for this prediction and calculation. A) The top view and side view of the bead model for the predicted Hik2 monomer. B) The hydrodynamic parameters of Hik2 calculated by US-SOMO modeler. The predicted $s_{20,w}$ value is indicated by a red box.

3.2.5.3 The sedimentation velocity experiment of Hik2

A Sedimentation Velocity (SV) experiment was performed, with the support of Ms. Liyana Azmi at University of Glasgow, to provide hydrodynamic information in order to give the shape and size information for Hik2 in solution, using absorbance optics. A series of protein concentration and different NaCl concentrations were used, including 1.8, 1.3, 0.7, 0.3 mg/ml of protein in low salt buffer (20 mM Tris-HCl, 10 mM NaCl (pH 7.5)), and 1.8, 1.0, 0.3 mg/ml of protein in high salt buffer (20 mM Tris-HCl, 500 mM NaCl (pH 7.5)). 5% glycerol was added to the buffers to prevent protein degradation when freezing and thawing.

Figure 3.19 shows the result of the SV experiment for 1.8 mg/ml of Hik2 in low salt buffer. The c(s) distribution was integrated to give the sedimentation coefficient (s), standardised to water at 20 ºC. There is one primary peak indicating that it is a predominantly homogenous sample. The value of $s_{20,w}$ at this protein concentration is 7.2, which is more than twice the predicted $s_{20,w}$ of 3.34 for Hik2 monomer, indicating that the protein complexes in solution are heavier and larger than the monomer, consistent with the previous result that Hik2 exists here as oligomers.

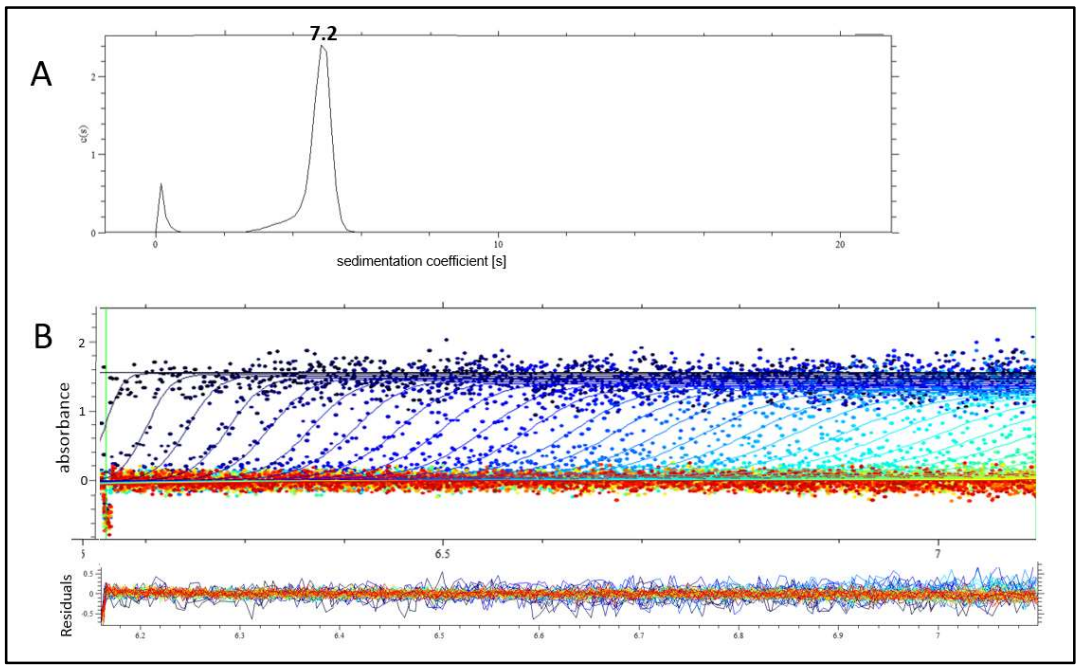
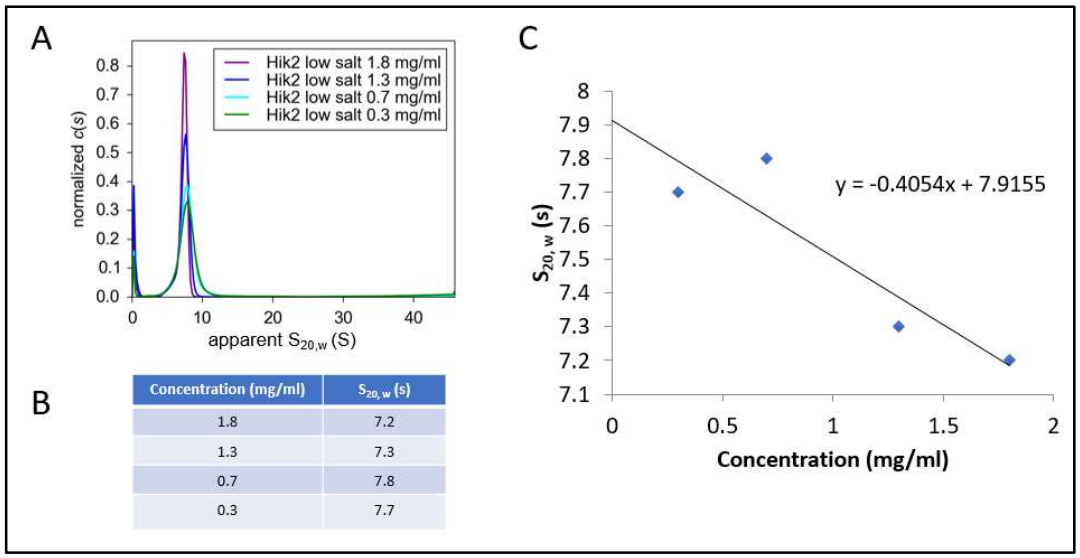


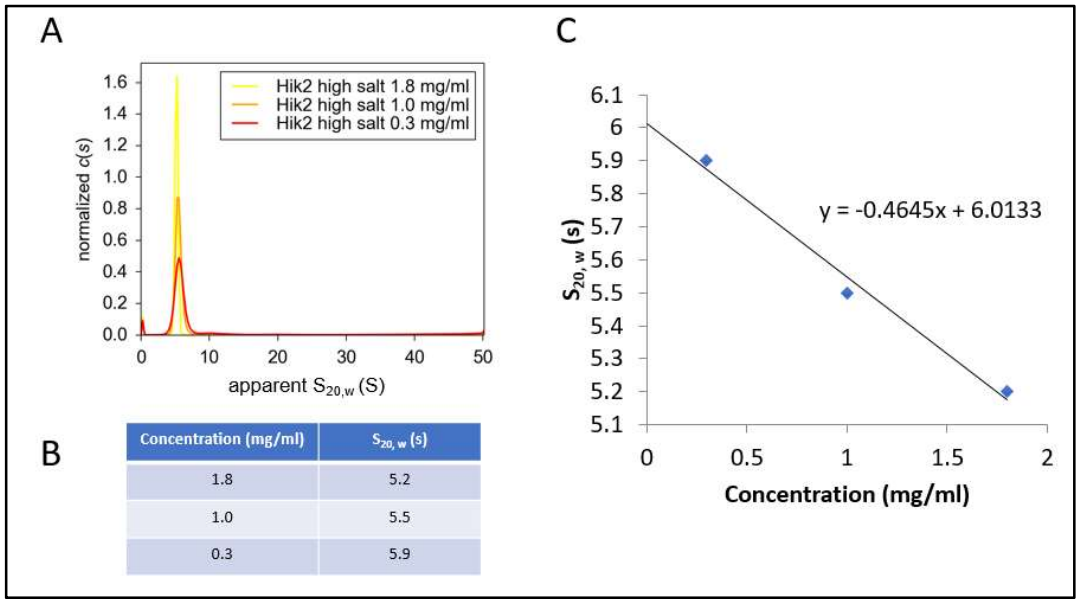
Figure 3.19 The SV experiment of Hik2 in low salt conditions at a concentration of 1.8 mg/ml. A) c(s) distribution derived via the continuous c(s) distribution model of SEDFIT, standardized to water at 20 ºC. A single peak is visible with an $s_{20,w}$ of 7.2 S. B) Statistical analysis. Superposition of experimental data (depicted as dots) and fitted (continuous line) SV profiles corrected for systematic noises. The data are observed to not fit well, with an RMSD of 0.119. The lower panel shows the superposition of the differences between the experimental and fitted curves.

After analysing all SV data, a single dominant peak in the c(s) distribution of Hik2 was found, under both low salt and high salt conditions. This indicated the protein was almost completely monodisperse in solution (Figure 3.20 A and Figure 3.21 A).



Concentration (mg/ml)	$S_{20,w}$ (s)
1.8	7.2
1.3	7.3
0.7	7.8
0.3	7.7

Figure 3.20 The sedimentation coefficient (s) for Hik2 in low salt conditions. A) The c(s) distributions at different protein concentrations from the SV of Hik2 in low salt conditions. B) The values of $s_{20,w}$ under different protein concentrations. C) The trendline obtained by plotting the values of $s_{20,w}$ against the loading protein concentration.



Concentration (mg/ml)	$S_{20,w}$ (s)
1.8	5.2
1.0	5.5
0.3	5.9

Figure 3.21 The sedimentation coefficient (s) for Hik2 in high salt conditions. A) The c(s) distributions at different protein concentrations from the SV of Hik2 under high salt conditions. B) The values of $s_{20,w}$ under different protein concentrations. C) The trendline obtained by plotting the values of $s_{20,w}$ against the loading protein concentration.

If there is one type of molecule in the solution, a standard sedimentation coefficient $s^{o}_{20,w}$ is usually reported, which can be obtained by the integrated values of $s_{20,w}$ against the loading concentration (Figure 3.20 C and Figure 3.21 C). The value of $s^{o}_{20,w}$ is a primary property of a molecule, which may change under different conditions, such as different pH or ionic strength (Campbell 2012). As can be calculated from Figure 3.20 C and Figure 3.21 C, the sedimentation coefficients of Hik2 at infinite concentration ($s^{o}_{20,w}$) in low salt conditions and for high salt conditions are determined to be 7.9 S and 6 S, respectively. Also, the slope of the trendline can indicate whether the protein self-associates. A positive slope suggests a self-associating protein and a decreasing trend shows a non-ideal protein. Both the trendlines in Figure 3.20 C and Figure 3.21 C are decreasing, indicating the non-ideality of Hik2.

3.2.5.4 The sedimentation equilibrium experiment of Hik2

A further experiment, Sedimentation Equilibrium (SE), was again used to probe the oligomeric state of Hik2. A series of Hik2 protein concentrations, at 1.8, 1.3, 0.7 mg/ml, in both low salt and high salt conditions, were tested, at speeds of 8,000 rpm and 12,000 rpm. Figure 3.22 displays a representative result which was done with 1.3 mg/ml Hik2 protein in low salt conditions at the speed of 12,000 rpm. The data were globally fitted using SEDPHAT software, however, it could not provide a good fit with the single species analysis model. Thus, a 2-species analysis model was applied, and a better fit with an RMSD of 0.006 was obtained. As boxed and highlighted in Figure 3.22, 84.7% of the sample is made up of proteins corresponding to a computed molecular mass of 181.96 kDa, while the remaining 15.3% fraction was 1020.613 kDa. The unusually large mass of the latter fraction might be aggregate, due to the long-time spin employed.

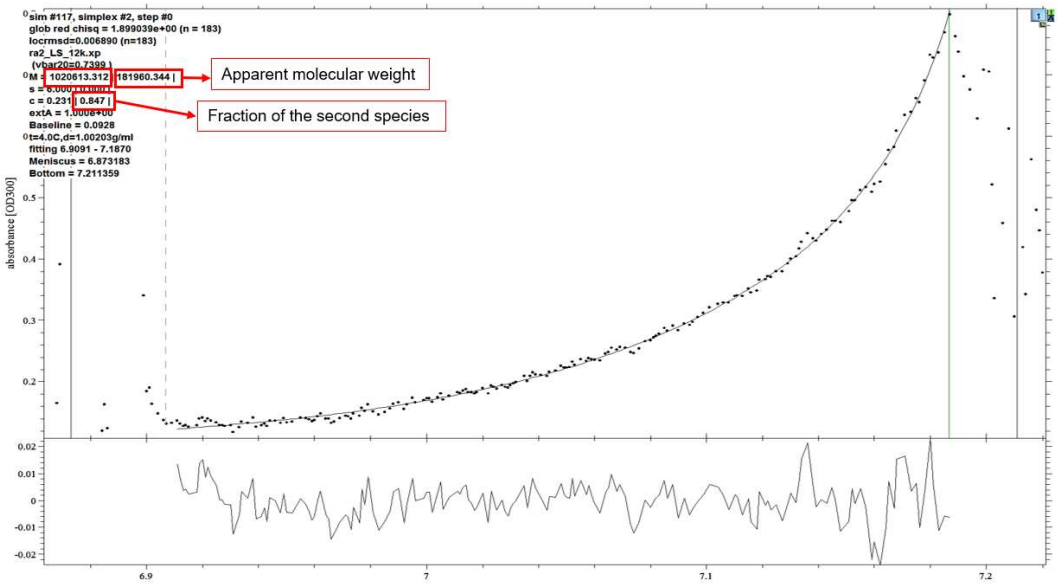


Figure 3.22 A representative result for the SE experiments. This particular run was performed at 4 ºC with 1.3 mg/ml Hik2 using low salt conditions, spun at 12,000 rpm. The SE data were globally fitted with a 2-species analysis model, resulting in a better statistical fit outcome (residuals inset) and well-fitted data (continuous lines), with an RMSD value of 0.006. The residual plot is shown below the graph. Red boxes indicate the apparent molecular mass (1020,613 kDa and 181.96 kDa) and the fraction of the second species (84.7%).

The computed SE data for Hik2 at different concentrations under low salt conditions at 12,000 rpm are shown in Figure 3.23 A. After plotting the majority of the computed mass from each concentration of Hik2 against the loading concentrations, the mass for Hik2 at infinite concentration can be calculated as 198.36 kDa (Figure 3.23 B), corresponding to the mass of a tetramer. The inconsistency with previous speculation is possibly due to the non-ideality of Hik2.

A

Concentration (mg/ml)	MW (kDa)	Percentage (%)
1.8	156.78	100
1.0	181.96	84.7
0.7	178.57	77.2

B

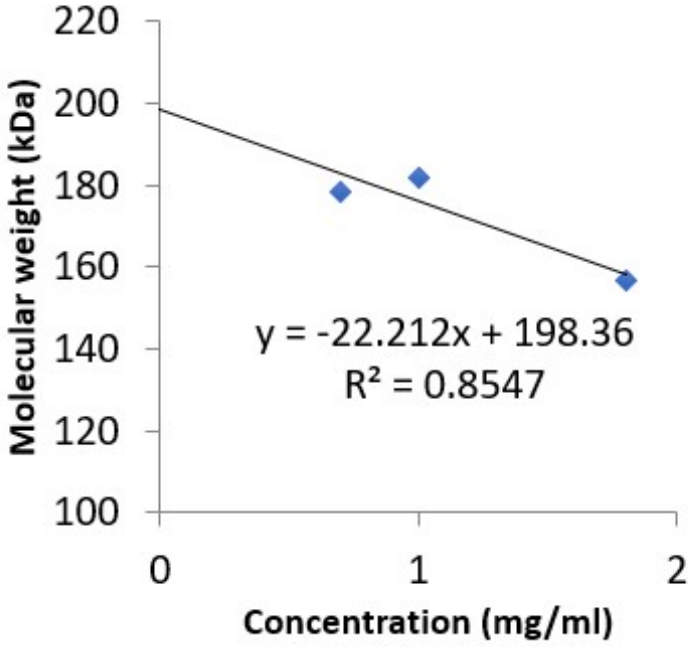


Figure 3.23 The SE data for Hik2 in low salt conditions. A) The computed molecular mass and percentage of Hik2 at different concentrations. B) The majority of the computed mass from each concentration of Hik2 is plotted against the loading concentrations. The SE experiment was performed at 12,000 rpm.

Unfortunately, the rest of the data were not analysable as the program failed to produce fits that could show a meaningful dataset.

3.3 Discussion and future work

The analysis of a protein sample that exists in various oligomeric states, in solution, is crucial for a complete characterisation of a protein under investigation, and for the functional study of said protein. Moreover, protein oligomerisation can prove significant as a trigger for various physiological pathways, or the cause of diseases, such as in Alzheimer's Disease. In this chapter, the oligomeric states of the full-length Hik2 and its subdomains were studied. Below, first, the novel properties of Hik2 protein oligomerisation observed are discussed.

3.3.1 The uncertainty of the apparent molecular mass of Hik2 oligomers in size exclusion chromatography

In size exclusion chromatography, proteins are separated not only by their mass, but their shape (rotational cross-section) also affects their migration. Since the calibration curve was calculated using the ideal globular standard proteins, and those only provide a general size indication, the determined molecular mass is likely to be inaccurate for those proteins with an elongated shape, and the latter will usually appear larger than their true molecular mass (Kirkwood et al. 2013). For the full-length *Synechocystis* Hik2, in the low salt conditions, the size exclusion chromatography result of Figure 3.7 A gave an apparent molecular mass of 337 kDa, which is close to, but larger than, that of a hexamer 291 kDa. This would indicate that the predominantly higher order oligomeric state of Hik2 is most likely a hexamer.

In such cases, for the non-sphere protein complexes, other methods should be used for molecular mass determination, such as AUC, (to be discussed in section 3.3.2) and

Small Angle X-ray Scattering (SAXS). Unfortunately, trials using SAXS failed due to the Hik2 protein aggregating while at the beam line itself. More work could, and should, be done to optimise the operating conditions for SAXS in the future.

3.3.2 The non-ideality of Hik2 observed in the AUC experiments

The equations used in AUC analysis are for single, ideal, non-associating solutes (Campbell 2012), thus if a protein is not acting ideally, the molecular mass calculated by the equations may prove to be inaccurate. The influence of molecular shape, such as for unfolded proteins or those of a more highly elongated shape, will make the protein complexes experience more friction from solvent, and tend thus to sediment more slowly. Therefore, the possibility of the highest order oligomer of Hik2 to be observed, that of a hexamer, should not be ruled out based on the SE result of AUC.

In addition, Figure 3.24 is an example of the SV data file of a typical multi-species protein system, displaying multiple boundaries in one scan (Brautigam et al. 2013). However, the AUC result in this thesis (Figure 3.19 B) shows only one boundary, meaning there is only one species in Hik2 sample under low salt conditions, which seems to be inconsistent with the consumption of the Hik2 equilibrium in solution. One possible reason why this may be being that this hexameric oligomeric form accounts for the majority of the protein, and the rest, i.e. tetramers, dimers, and monomers, are too small for the AUC to technically detect. Another reason may be that, according to the previous results, the maximum component in the equilibrium of Hik2 solution are indeed hexamers, the most stable oligomeric state. During the sedimentation experiments, the faster sedimenting complexes, the hexamers, must migrate through a

solution of the slower sedimenting components, such as the tetramers, dimers, and monomers, which may also readily re-associate to form hexamers during the migration.

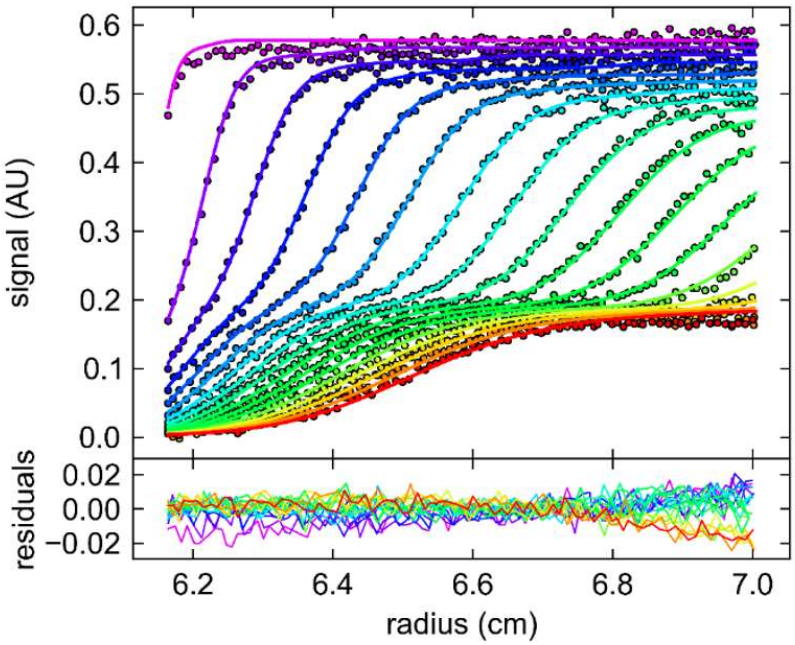


Figure 3.24 A typical SV analysis data file of a multi-species protein sample. Multiple boundaries are observed. Figure from (Brautigam et al. 2013).

3.3.3 The salt endurance of Hik2

A previous study showed that the autophosphorylation activity of Hik2 decreases when increasing the concentration of NaCl in buffer, and this activity is almost entirely inhibited with 1 M of NaCl in the buffer (Ibrahim et al. 2016). In this chapter, it was shown that NaCl dissociates the higher order oligomers of Hik2 (see section 3.2.3). Moreover, Hik2 was eluted from Ni^{2+} affinity chromatography column in the buffer containing 300 mM NaCl (section 2.3.1), and was observed to be able to form higher order oligomers again when decreasing NaCl concentration to 10 mM during size exclusion chromatography (section 3.2.2), and showed the highest activity in the autophosphorylation assay after being desalted into the buffer containing only 10 mM

Tris-HCl (Ibrahim et al. 2016). Thus, the effect of NaCl on both oligomeric state transduction and autokinase activity must be reversible.

Ammonium acetate (CH_3COONH_4) has been widely used to buffer proteins for mass spectrometry due to its volatility at low pressure, to minimise the interference of salt crystals and other mass spectral peaks caused by the buffer (Berman et al. 2008). 10 to 500 mM of ammonium acetate proved to be applicable for standard proteins, such as chicken egg white lysozyme, bovine pancreas trypsin, bovine β-lactoglobulin A, amongst others (Gavriilidou et al. 2015). Interestingly, during the mass spectrometry experiment in section 3.2.2, it was observed that Hik2 could not cope with as little as 100 mM of ammonium acetate, since it aggregated and precipitated immediately when ammonium acetate was added to the buffer, this process proving irreversible even after ammonium acetate was removed from the buffer. Another soluble histidine kinase, ExsG, can stand as much as 800 mM of ammonium acetate (Wojnowska et al. 2013). The unusual sensitivity to ammonium acetate likely indicates structural or surface charge differences for Hik2 compared to other soluble histidine kinases.

3.3.4 The function of the DHp domain in Hik2

In this thesis, both results of size exclusion chromatography and native-PAGE show that all Hik2 protein variants that contain the DHp domain, such as the full-length Hik2, the kinase domain of Hik2, as well as DHp domain alone, appear to form higher order oligomers.

The structure of the homodimeric DHp domain is comprised of a closed, four-helical bundle with a left-handed twist, formed by two identical alpha-hairpin subunits. The

HisKA domain subfamily is usually seen as the dimerisation-effecting DHp domain in most HKs, and it also directly contacts the cognate response regulators (Willett and Kirby 2012). There is crystal structure information for DHp domains, but most of the proteins have been crystallised together with other domains, such as the HAMP and CA domains, e.g. the crystal structure of the DHp domain with HAMP domain in EnvZ (Ferris et al. 2014), and the crystal structure of the DHp domain with CA domain in CckA (Dubey et al. 2016). There is currently no DHp domain, alone, having been reported as being crystallised.

The structure of the DHp domain of *Synechocystis* Hik2 was predicted in this study using the online server Phyre2 (Kelley et al. 2015). Compared with the kinase domain predicted by I-TASSER, the DHp domain structure matches well, consisting of two α-helices, connected by a short loop (Figure 3.25).

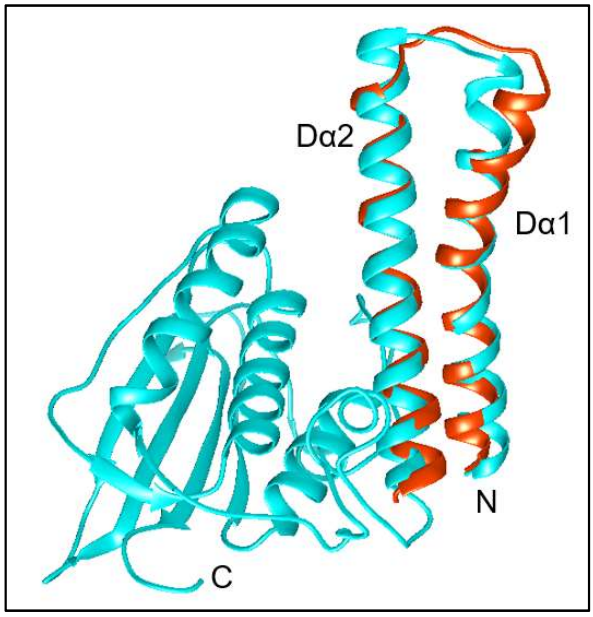


Figure 3.25 The predicted structures of Hik2 DHp domain and kinase domain. The orange one represents the DHp domain predicted by Phyre2, while the cyan one represents the kinase domain predicted by I-TASSER. Figure was prepared using UCSF Chimera (Pettersen et al. 2004).

The DHp domain has an all α-helices structure, similar to the all α-helix type of certain extracytosolic sensor domains. For example, the NarX protein in *E. coli* has an all α-helices sensor domain, also the TorS in *Vibrio parahaemolyticus* has double helical domains (section 1.1.1.3 in Chapter 1). The NarX sensor forms a dimer, and binds a nitrate ion partially buried between the two α-helices in the interface (Cheung and Hendrickson 2009). Combined with the evidence that Na^+ ions being proved to inhibit autophosphorylation activity of Hik2 (Ibrahim et al. 2016), and NaCl regulate the oligomeric states of the Hik2 DHp domain, and furthermore the full-length and truncated Hik2 protein, it is likely that the DHp domain in Hik2 binds directly to Na^+ ions, responding to this signal.

In a histidine kinase, most of the DHp domains are responsible for the dimerisation and phosphor-accepting function while the sensor domains sense signals. However, interestingly, it was reported that one transmembrane histidine kinase, EnvZ, senses an osmolar signal via its DHp domain, regulating expression of outer membrane proteins (Wang et al. 2012). High osmolality, such as 0.5 M NaCl or 20% (w/v) sucrose, can alter the conformation of the DHp domain and increase the autophosphorylation of EnvZ, and the phosphoreceptor histidine His^{243} is required for this osmosensing. The authors proposed a stretch-relaxation model of the DHp domain, that weaker intrahelical H-bonds exist at low osmolality, while high osmolality increases the stabilisation of intrahelical H-bonds, making the conformation more folded (Wang et al. 2012). Comparing with EnvZ, in this study, Na^+ may interfere with salt bridges that stabilise interunit interfaces of Hik2, and induce a conformational change of the DHp domain, leading to disruption of interunit interactions in the hexamers and favouring the formation of the protein-protein interface of an inactive tetramer.

3.3.5 The higher order oligomeric state of Hik2

Most histidine kinases appear to exist the homodimeric form, including the classic transmembrane histidine kinases (Gao and Stock 2009), such as HK853 (Marina et al. 2005), DesK (Albanesi et al. 2009), HK853 (Casino et al. 2009), and also cytoplasmic ones, such as BphP1(Li et al. 2010).

However, it has been reported that an HWE family protein ExsG, a hybrid histidine kinase containing an N-terminal receiver domain, forms an active hexamer in solution via its HWE HK core, corresponding to the DHp domain (Wojnowska et al. 2013). It is possible that this kind of hybrid histidine kinase, with both receiver domain and kinase domain, can relay the signal between an upstream histidine kinase and downstream response regulator, acting simultaneously as a response regulator and a histidine kinase. Furthermore, it has been suggested that ExsG and three other proteins that encode with it, a histidine kinase photoreceptor BphP1, and the stand-alone receiver domain proteins AgR and ExsF, function in a two-component-like signalling pathway. The signalling initiated from BphP1 are branched, and transferred partially to AgR directly or to ExsF via the mediator ExsG. The hexameric ExsG exists in equilibrium between the active 'open/relaxed' and inactive 'closed/compact' conformations, regulated by the phosphorylation state of its receiver domain (Wojnowska et al. 2013).

It has been introduced in the review (Allakhverdiev and Murata 2008) that salt stress induces the reversibly rapid decrease and irreversibly slow decrease of the photosynthetic activities of the two photosystems, especially PS II. It has been shown that in the presence of high concentration of salt (0.5 M NaCl), the transcription and translation of the *psbA* genes (which encode the D1 protein) were suppressed, and the

repair of PS II was inhibited in *Synechocystis* (Allakhverdiev et al. 2002). In the meanwhile, the de novo synthesis of D1 protein in *Synechococcus* was also inhibited by salt stress, leading to the decrease of the PS II activity (Ohnishi and Murata 2006). qRT-PCR analysis indicated that the phycobilisome linker polypeptide (ORF2155) was down-regulated by 3-fold under salt stress in *Arthrospira* (Wang et al. 2013). For the Hik2 protein discussed in this thesis, both from *Synechocystis* and *Thermosynechococcus*, a model of the signal transduction mechanism is proposed in Figure 3.26 (Ibrahim et al. 2017). In the absence of Na^+, Hik2 exists in equilibrium between the active hexameric, tetrameric and monomeric forms, and the hexameric Hik2 is autokinase-active and transfers phosphoryl groups to its cognate response regulators Rre1 and RppA, activating the gene expression of the light-harvesting phycobiliproteins; and the phospho-Rre1 inhibits the transcription of salt-tolerance genes. However, when elevating the concentration of Na^+, the active hexameric form of Hik2 dissociates into an inactive tetrameric form, so that the Rre1 and RppA remain in their unphosphorylated state, which cannot active the photosystem genes; meanwhile the expression of salt-tolerance genes is no longer repressed, allowing the organism to react to the salt stress rapidly.

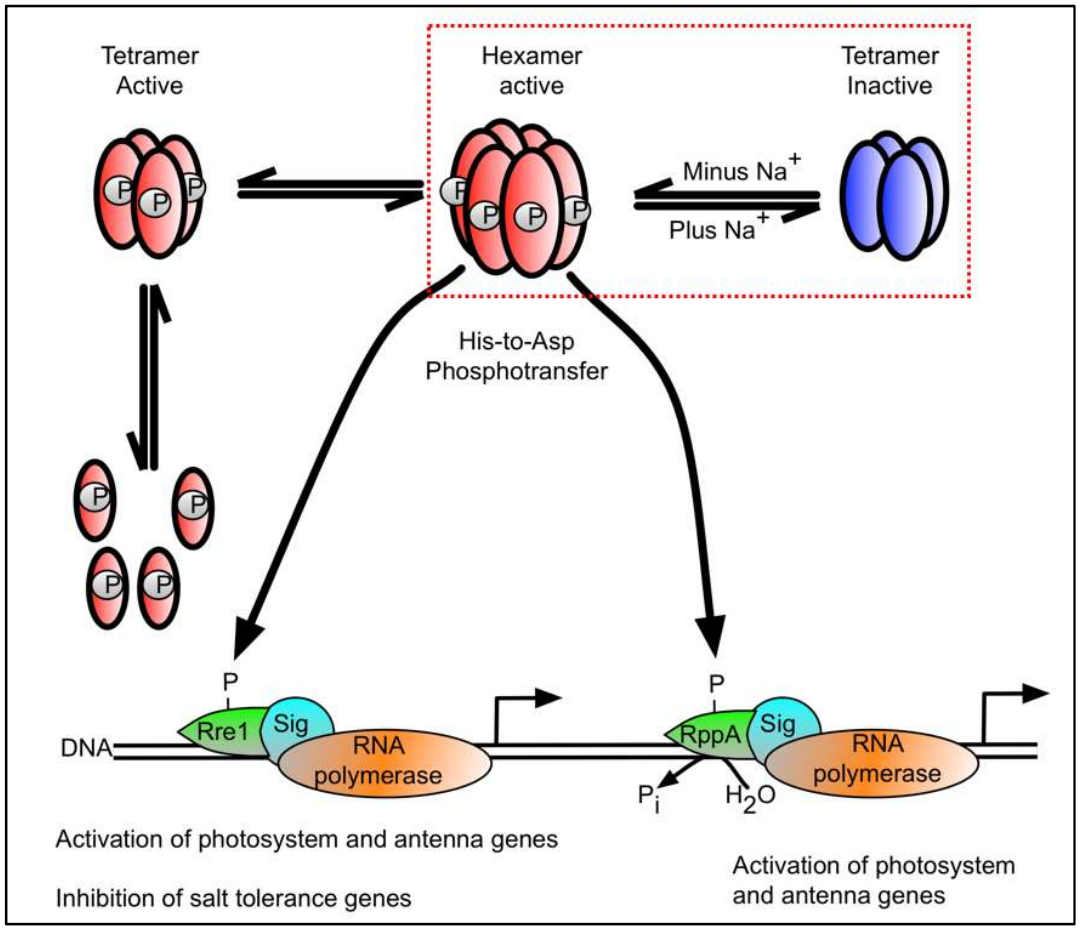


Figure 3.26 A proposed model for a Hik2 signal transduction pathway in cyanobacteria based on its oligomeric state. The hexameric, tetrameric and monomeric form of Hik2 with circled P (the phosphoryl group) are autokinase-active, while the tetrameric form without circled P is autokinase-inactive, and the oligomeric state of Hik2 is regulated by Na^+. The square with red dotted line indicates the hypothesis/results of the work in this chapter that the active hexameric form of Hik2 is converted to an inactive tetramer upon Na^+ stress. Figure modified from (Ibrahim et al. 2017).

Although both are reported to exist in the active hexameric form, Hik2 and ExsG have significant differences. ExsG is proposed to be both active and inactive as a hexamer, in an equilibrium in solution, while hexameric Hik2 is proposed to be active form, and this becomes an inactive tetramer when converted by the presence of Na^+.

To further study the interaction and activation between Hik2 and Rre1, a phosphotransfer analysis could be performed with chemical cross-linking in order to identify the oligomeric form of Hik2 that is able to transfer its phosphoryl group to Rre1.

3.4 Conclusions

It is well known that protein structure is related to its function. It was observed in this chapter that with full-length Hik2, the DHp domain alone and the kinase domain exist as oligomers in aqueous solutions of low salt, from the results of a variety of methods above, including size exclusion chromatography, chemical cross-linking, and dynamic light scattering. This outcome is not common for histidine kinases in general. The experiments also showed that adding NaCl resulted in a dissociation of the higher order oligomers. Considering this, together with the autophosphorylation activity inhibited by Na^+ ions (Ibrahim et al. 2016), it can be assumed that the Na^+ ions bind the DHp domain of Hik2, and are thus related to oligomerisation of the protein and its autophosphorylation activity. Moreover, the native-PAGE profiles and AUC results indicate there might be different conformations of Hik2 oligomers.

Chapter 4 Structural characterisation of the Hik2 and CSK sensor domains

4.1 Overview

The sensor domain of any histidine kinase is considered to be the most important domain, due to the sequence and structure differences present between sensor domains of species, which allow histidine kinases to react to different signals and possess unique functions. Thus, in these studies, more attention was paid to the sensor domains of Hik2 and CSK. There is currently no structural information for the sensor domain of CSK, named here as the Ligand Binding Domain (LBD), although CSK may sense a particular redox signal (see Chapter 1, section 1.3.3). The sensor domain of Hik2 has been identified as a GAF domain, however, even though many crystal structures and functions of other GAF domains have been revealed (see Chapter 1, section 1.1.1.3), the signalling mechanism of Hik2's GAF domain is still unknown. In this chapter, the sensor domains of *Synechocystis* Hik2 and *Arabidopsis* CSK were studied further with the expectation of obtaining a structure from X-ray diffraction analyses or by using Nuclear Magnetic Resonance (NMR) spectroscopy.

4.2 Results

In the following sections, the sensor domains of *Synechocystis* Hik2 (GAF domain) and *Arabidopsis* CSK (LBD) were analysed, and attempts were made to purify and solubilise them, such as *in vitro* denaturation and refolding, and by ligating the two sensor domains with MBP or SUMO tags. Also, additional tertiary structural information for *Synechocystis* Hik2 was obtained by limited proteolysis.

4.2.1 Sequence analysis of the sensor domains of Hik2 and CSK

4.2.1.1 The sensor domain of Hik2 is a GAF domain

Sequence analysis using the SMART online server (Schultz et al. 1998, Letunic et al. 2015) indicates that the sensor domain of Hik2 is a GAF domain, a terminology derived from its presence in cyclic GMP (cGMP)-regulated cyclic nucleotide phosphodiesterases, certain Adenylyl cyclases, and the bacterial transcription factor FhlA. The GAF domain also exists in many other signalling and sensor proteins (see Chapter 1, section 1.1.1.3).

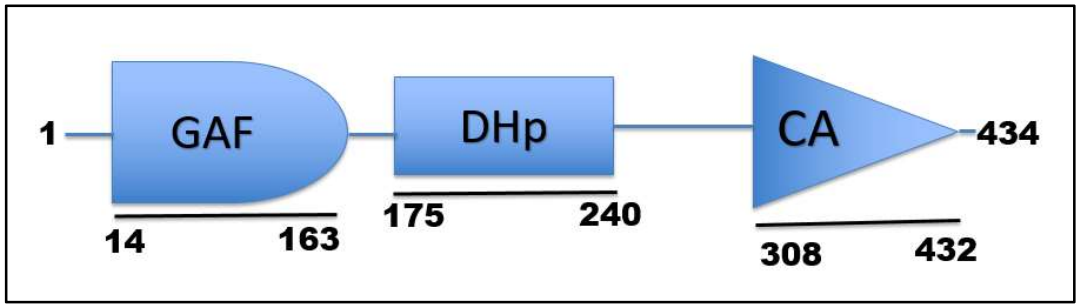


Figure 4.1 Domain prediction of *Synechocystis* Hik2 carried out using the SMART online server. Two key domains were most prominently revealed: the DHp domain (for structural dimerisation phosphor-accepting function) and CA domain (for catalytic function).

The PAS domain is so named after the first three discovered proteins, those being the Period circadian protein, Aryl hydrocarbon receptor nuclear translocator protein and the Single-minded protein, and this PAS domain shares significant similarity with the GAF domain (see Chapter 1, section 1.1.1.3).

4.2.1.2 There is no known structure identical to the sensor domain of CSK

When analysing the whole CSK sequence for *Arabidopsis thaliana* using SMART online server (Schultz et al. 1998, Letunic et al. 2015), no other known structure for its sensor domain was reported, as yet.

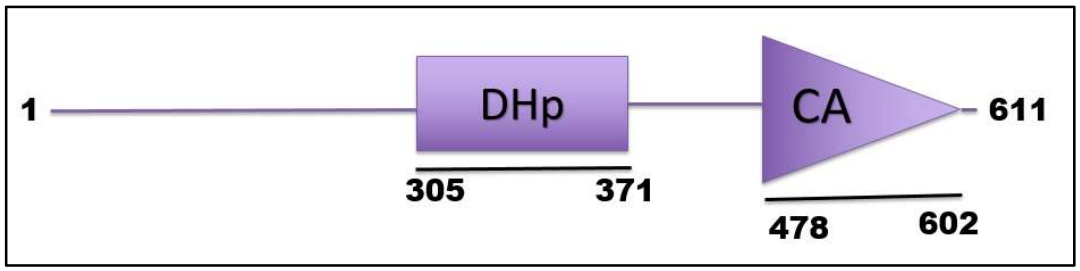


Figure 4.2 Domain prediction of *Arabidopsis* CSK carried out using the SMART online server. Two key domains were most prominently revealed; the DHp domain (for structural dimerisation phosphor-accepting function) and CA domain (for catalytic function).

4.2.2 Production and purification of the Hik2 and CSK sensor domains

4.2.2.1 The sensor domains were insoluble when expressed at various temperatures

The GAF domain of *Synechocystis* Hik2 (Hik2_GAF) was cloned into a pET_14b vector, expressed and purified under different conditions, as described in the following section. However, when incubated at 30 ºC for 3 h or 18 ºC for 16 h, the Hik2_GAF protein was expressed, but found to be insoluble (Figure 4.3 lanes 3, 5, 8, 10).

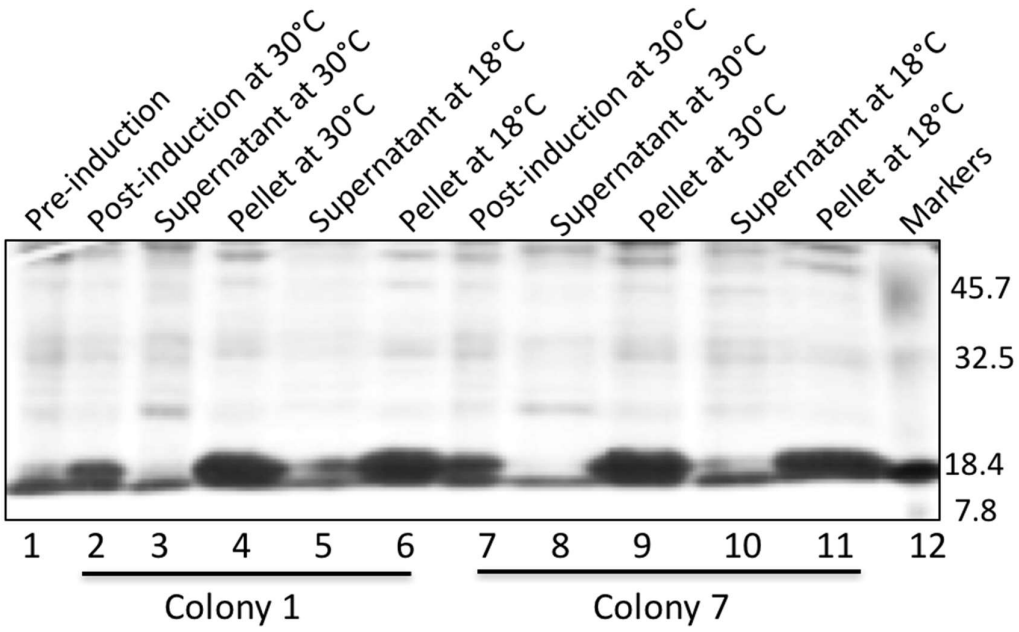


Figure 4.3 SDS-PAGE profile of the isolated Hik2_GAF domain expressed at different temperature. Lanes 1 to 11, as labelled; lane 12, protein molecular mass markers. The SDS-PAGE was done with two different single colonies, expressed at 30 ºC and 18 ºC. The kDa values are shown on the right.

4.2.2.2 The sensor domains are insoluble even under different conditions

To improve the solubility of the Hik2_GAF protein, different lysis buffers were used to suspend the pellet in, either at different pH values, ranging from the acidic pH 5.7, through to basic pH 8.5 (Figure 4.4 A), or with higher concentration of salt and DTT (Figure 4.4 B), however, the protein remained stubbornly insoluble. This suggested that the protein was most likely present within inclusion bodies.

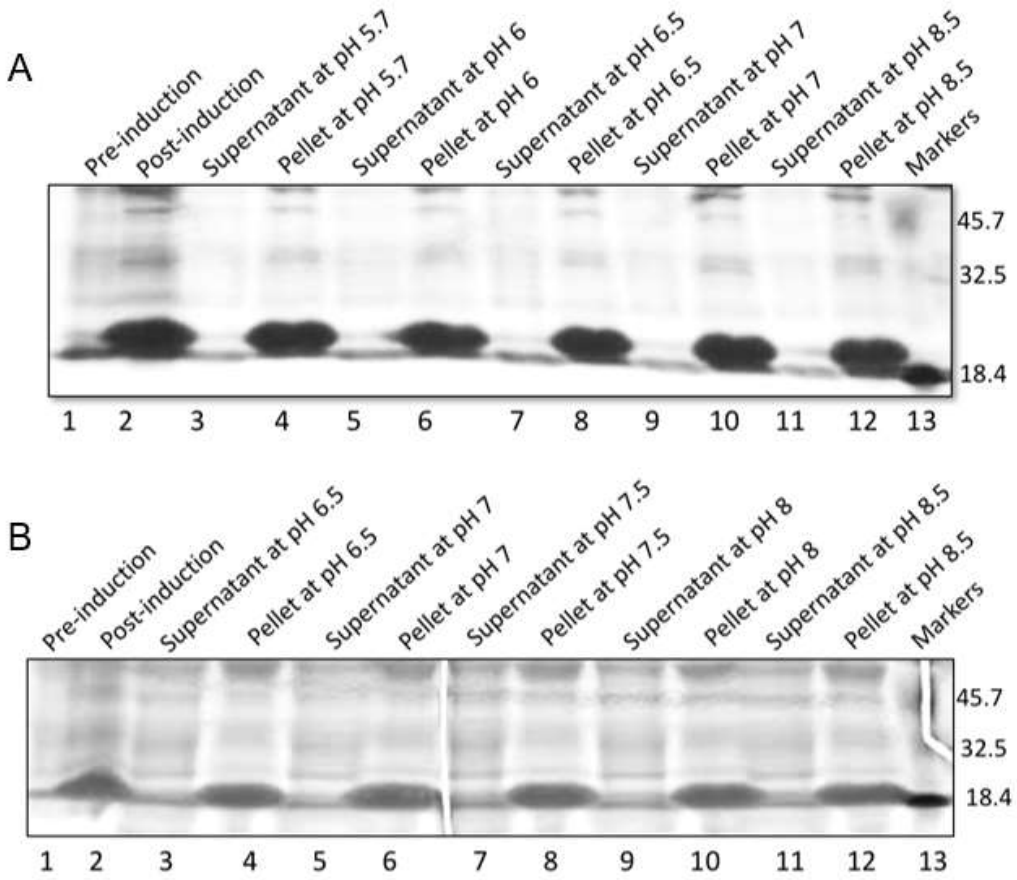


Figure 4.4 SDS-PAGE profiles of Hik2_GAF lysis samples in buffers of differing pH. A) Hik2_GAF samples expressed at 18 ºC for 16 h and resuspended in lysis buffer (20 mM Phosphate/Tris-HCl, 300 mM NaCl) at pH 5.7, 6, 6.5, 7, 8.5. Lanes 1 to 12, as labelled; lane 13, protein molecular mass markers. The kDa values are shown on the right. B) Hik2_GAF samples expressed at 18 ºC for 16 h and resuspended in high salt buffer with reducing agent (20 mM Phosphate/Tris-HCl, 500 mM NaCl, 20 mM DTT) at pH 6.5, 7, 7.5, 8, 8.5. Lanes 1 to 12, as labelled; lane 13, protein molecular mass markers. The kDa values are shown on the right.

A protein folding prediction with the CSK sequence for *Arabidopsis thaliana* was performed using FoldIndex (Prilusky et al. 2005), and amino acids 51 to 207 (CSK_LBD) were chosen to be commercially synthesised (Eurofins Genomics Ltd, Ebensburg, Germany) and cloned into a pET_14b vector. Even from this prediction

based result, solubility was found to be similar to that obtained in Figure 4.4 (data not shown), and essentially it was decided to move forward towards attempting the solubilisation in the following denaturation and refolding studies below.

4.2.3 Studies to improve the solubility of the sensor domains

4.2.3.1 *In vitro* denaturation and refolding

The pellet for each pellet sample, assumed to contain inclusion bodies, was isolated and solubilised by 6 M Guanidine-HCl, then applied to a commercial Ni^{2+}-affinity bead batch test, and refolded with buffer lacking Guanidine-HCl (Figure 4.5). As is clear in lanes 5 and 11, a very small amount of the protein appeared to have refolded.

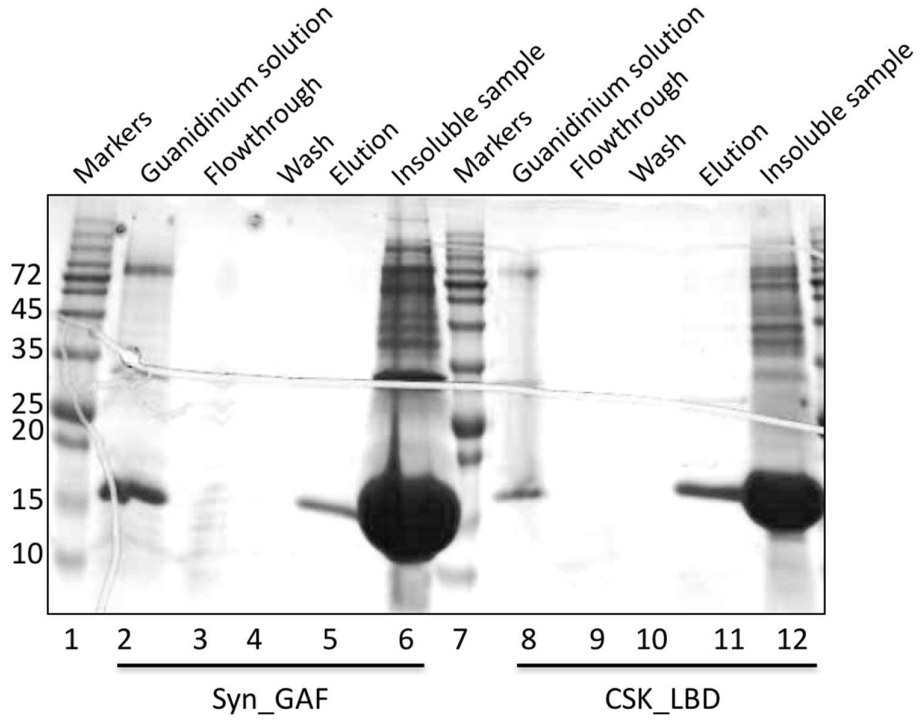


Figure 4.5 SDS-PAGE profile of samples from a Ni^{2+}-affinity denaturation and refolding batch test on isolated Hik2_GAF and CSK_LBD. Lanes 1 and 7, protein molecular mass markers; lanes 2 to 6 and 8 to 12, as labelled. The kDa values for the molecular mass markers are shown on the left.

Then each protein domain was expressed at a larger scale, in terms of now a 1 L culture. The inclusion bodies were isolated, washed by 2 M urea, and solubilised with 6 M Guanidine-HCl, then applied to Ni^{2+}-affinity column. The buffer was changed to 6 M urea to wash the column, then the denatured protein was eluted with 6 M urea (Figure 4.6 lane 4), and the refolded protein was then eluted either after decreasing the urea concentration rapidly (Figure 4.6 lane 5) or slowly (Figure 4.6 lane 6). In this way, around 0.3 mg of refolded Hik2_GAF could be obtained from 1 L of *E. coli* BL21 culture.

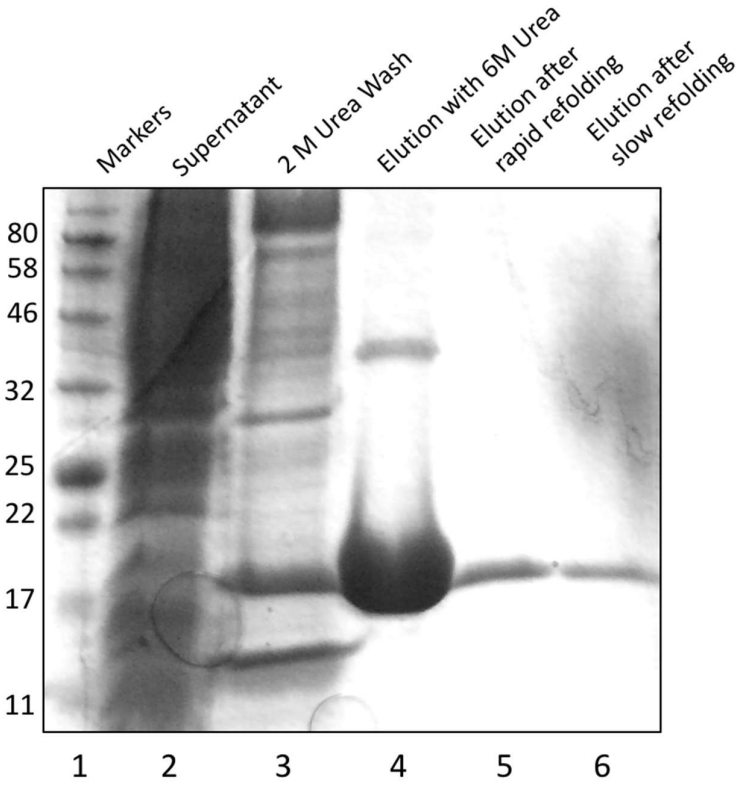


Figure 4.6 SDS-PAGE profile of denaturation and refolding of the Hik2_GAF protein. Lane 1, protein molecular mass markers; lanes 2 to 6, as labelled. The kDa values for the molecular mass markers are shown on the left.

4.2.3.2 Protein expression of the MBP and SUMO fusion partners, and subsequent 3C protease cleavage

It was reported that the MBP (Maltose-Binding Protein) and SUMO (Small Ubiquitin-related Modifier) fusion partners may enhance protein expression and solubility, and also protect target proteins from degradation. Both Hik2_GAF and CSK_LBD sequences were cloned into pOPINM and pOPINS3C vectors. The pOPINM vector carries a cleavable N-terminal His-MBP tag, and the pOPINS3C vector carries a cleavable N-terminal His-SUMO tag. These fusion vectors were transformed into BL21 *E. coli* cells and induced by 0.8 mM IPTG at 18 ºC for 16 h. The expressed protein was then applied to a Ni^{2+}-affinity column. As observed in Figure 4.7, lane 4 and lane 7, the Hik2_GAF_MBP fusion (60 kDa) and CSK_LBD_MBP fusion (60.5 kDa) partners were expressed (highlighted in red boxes). However, the elution was observed to be impure, with two apparent contamination bands present in the profile at ~ 70 kDa and 25 kDa. There are also two bands around 43 kDa and 17 kDa, which might be the degraded products MBP alone (43.5 kDa) and Hik2_GAF (16.6 kDa). However, most of the Hik2_GAF_SUMO fusion protein (29.8 kDa) and CSK_LBD_SUMO (30.3 kDa) fusion protein appeared in the pellet after sonication (data not shown) and were not seen in the SDS-PAGE profile (Figure 4.7 lanes 10 and 13), thus it appears unlikely that the SUMO tag improved the solubility of the Hik2 and CSK sensor domain protein complexes.

Subsequently, the buffer was exchanged into 20 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1 mM DTT (pH 7.8), either by PD-10 column or by dialysis, in order to perform cleavage by 3C protease. From Figure 4.7, lanes 2, 3 and lanes 5, 6, it should be noted

that the MBP tag and target protein were indeed cleaved, but this cleavage appeared incomplete.

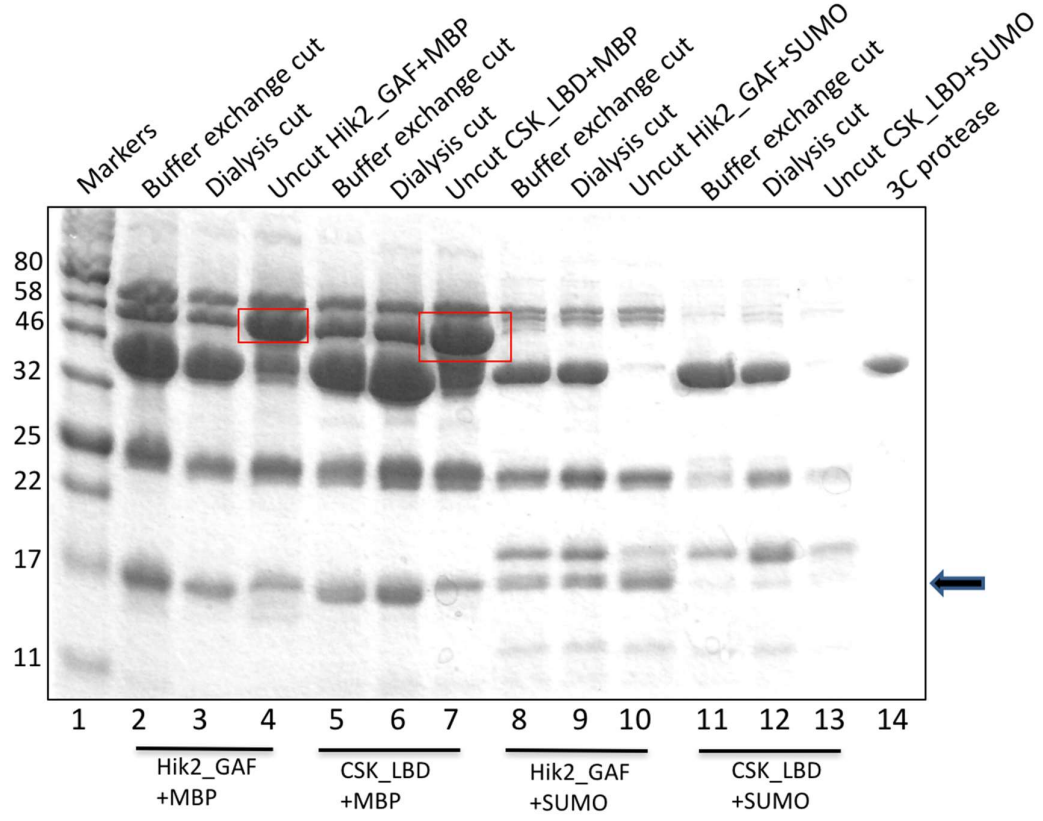


Figure 4.7 SDS-PAGE profile of MBP and SUMO fusion proteins expression and the cleavage by 3C protease. Lane 1, protein molecular mass markers; lanes 2 to 6, as labelled. The kDa values for the molecular mass markers are shown on the left. Red boxes denote the MBP with Hik2_GAF and CSK_LBD fusion partners, respectively. The arrow denotes the protein band at ~ 16 kDa, attributable to the cleaved Hik2_GAF domain.

Following this, an amylose resin column, which specifically binds MBP-tagged proteins, was used to purify the Hik2_GAF+MBP and CSK_LBD+MBP fusion protein complexes (Figure 4.8). The purity of fusion proteins was found to have improved, and less degradation was also apparent.

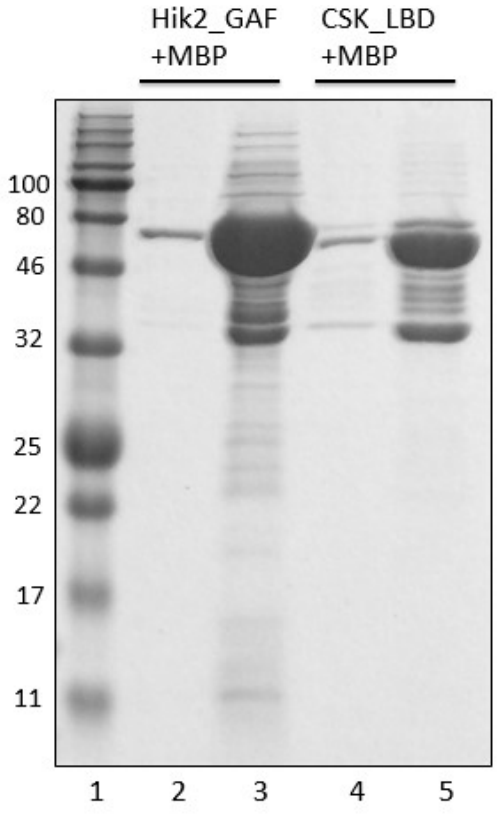


Figure 4.8 SDS-PAGE profile of Hik2_GAF+MBP and CSK_LBD+MBP fusion protein complexes, as purified by amylose resin column. Lane 1, protein molecular mass markers; lanes 2 and 3, elutions 1 and 2 of Hik2_GAF+MBP; lanes 4 and 5, elutions 1 and 2 of CSK_LBD +MBP. The kDa values for the molecular mass markers are shown on the left.

Subsequently, the elution proteins were concentrated and buffer exchanged into 20 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1 mM DTT (pH 7.8) by Vivaspin column (10 kDa), in order to perform cleavage by 3C protease. As depicted in Figure 4.9, lanes 3, 4 and 6, 7 shows that parts of the fusion proteins were cut apart, but cleavage was insufficient.

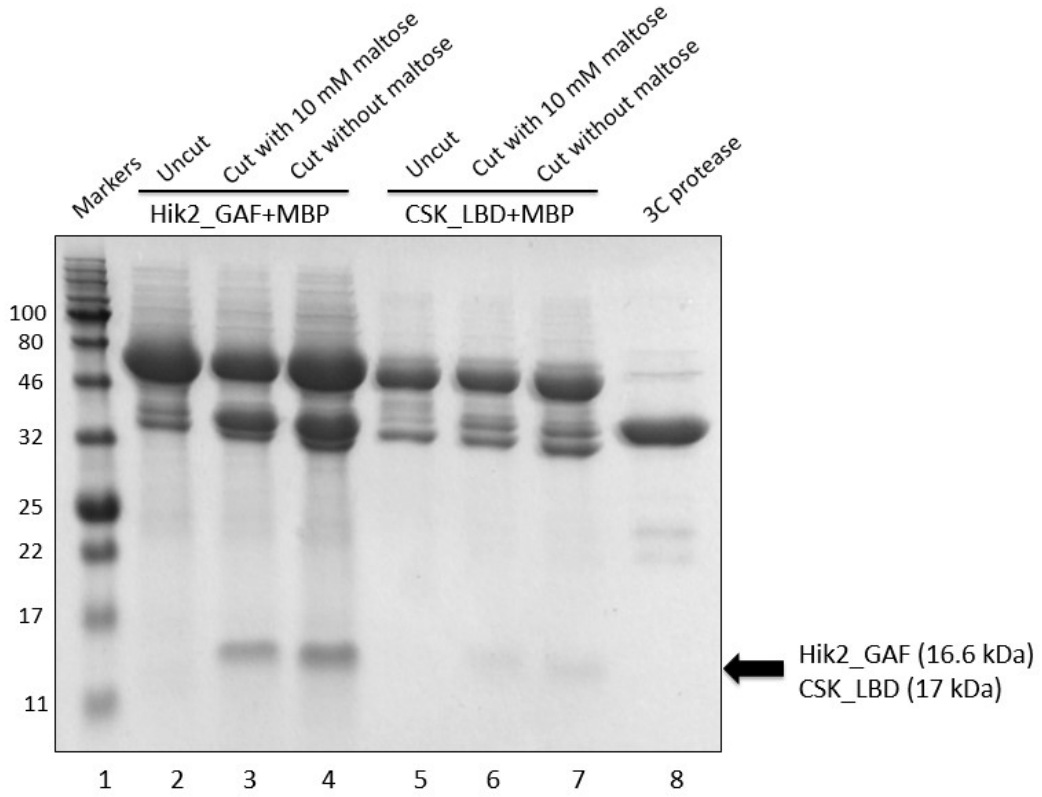


Figure 4.9 SDS-PAGE profile of the 3C protease cleavage of Hik2_GAF+MBP and CSK_LBD+MBP fusion protein complexes. Lane 1, protein molecular mass markers; lanes 2 to 8, as labelled. The kDa values for the molecular mass markers are shown on the left. The arrow emphasises protein bands at ~ 16.6 and 17 kDa for the cleaved Hik2_GAF and CSK_LBD proteins, respectively.

4.2.4 Structure prediction of CSK and Hik2 sensor domains

Since it proved difficult to purify the CSK and Hik2 sensor domain proteins in high enough quality and quantity, protein modelling was applied to predict their structures from their amino acid sequences. Several software packages and online servers can be used to predict protein tertiary structure, such as Phyre2 (Kelley et al. 2015), ESyPred3D (Lambert et al. 2002), SWISS-MODEL (Biasini et al. 2014), I-TASSER (Zhang 2008), etc.

Most software builds 3D models using homology detection methods, however, the homologous sequences of the CSK_LBD were difficult to find, and the coverage of the predicted structure was very low for these servers such as SWISS-MODEL (Biasini et al. 2014) and Phyre2 (Kelley et al. 2015), thus I-TASSER, a technique that is based on the secondary-structure enhanced Profile-Profile threading Alignment (Zhang 2008), was used to build the model of CSK_LBD. The predicted structures are displayed in Figure 4.10.

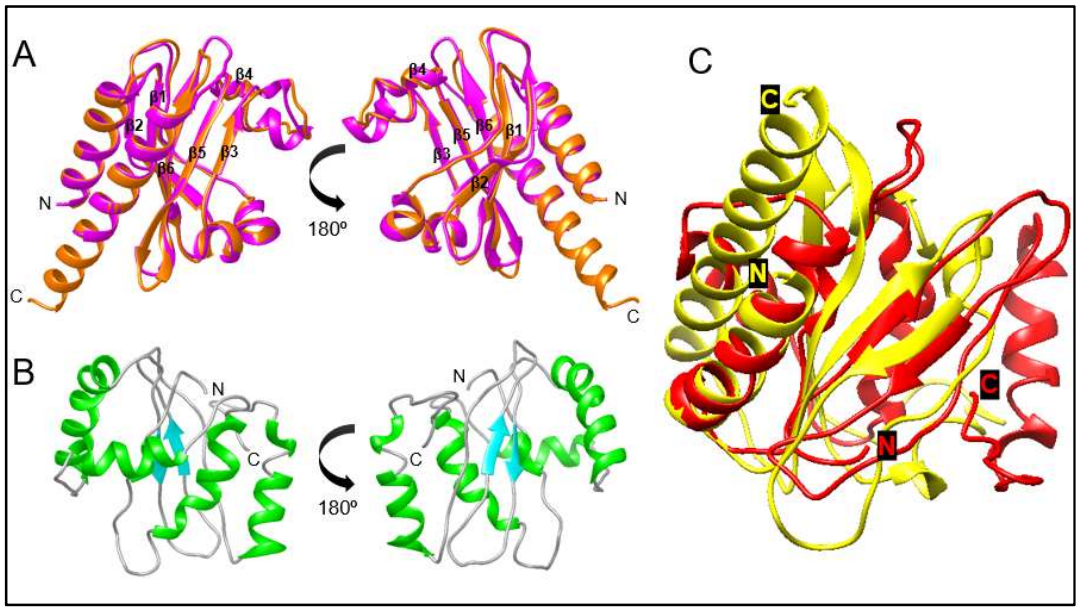


Figure 4.10 The predicted structures of Hik2 and CSK sensor domains. A) The comparison of the predicted Hik2_GAF structure, by I-TASSER (coloured in orange) and that by Phyre2 (coloured in purple). B) The predicted CSK_LBD structure by I-TASSER, α-helices are rendered in green, and β-strands in cyan. C) The structure comparison of the I-TASSER predicted Hik2_GAF (coloured in yellow) and CSK_LBD (coloured in red). Both of the two predicted sensor domains are mixed α/β fold structure. However, the Hik2_GAF is well-folded, possessing more β-sheets in the centre, while the CSK_LBD is more relaxed, with many random coils and four α-helices flanking around a small β-sheets core. This figure was drawn using UCSF Chimera (Pettersen et al. 2004).

The structure of Hik2_GAF was predicted by both I-TASSER and Phyre2 (Figure 4.10 A), according to sequence alignment and homology with the blue-light absorbing form of GAF domain from cyanobacteriochrome PixJ in *Thermosynechococcus elongatus*, PDB code 4FOF (Burgie et al. 2013). It may be seen, from the comparison, that the two predicted structures using different software are very similar with backbone RMSD of 1.13 Å (Maiti et al. 2004). The predicted Hik2_GAF has a typical GAF domain structure with a six-stranded antiparallel β-sheet core, and two α-helices flanking on each side, forming the outer layer with short loops. The predicted CSK_LBD structure is also a mixed α/β fold with only two small β-sheets in the centre and four α-helices surrounding them, containing many disordered flexible loops (Figure 4.10 B). The structure comparison between the two sensor domains shows that they have similar overall structure, a mixed α/β fold, possessing similar β-sheet core and a few α-helices with nearly-identical directions, but the CSK_LBD is relatively loose while the Hik2_GAF more compact (Figure 4.10 C).

4.2.5 Studies on other domains of Hik2

Limited proteolysis may be used to investigate the conformational information of a protein such as its folding. The secondary structure loops present are more vulnerable than a folded domain because the cleavage sites are more exposed. This technique was subjected to the isolated Hik2 protein in order to find any stable fragments suitable for gaining crystal forms prior to optimising such conditions or, ideally, diffraction studies. Time courses were taken using trypsin as the protease (Figure 4.11). A stable fragment with a molecular mass of ~ 28 kDa was identified that remained even after 1 h of incubation with trypsin, at 22 ºC, and also after an overnight incubation, with trypsin, at 4 ºC.

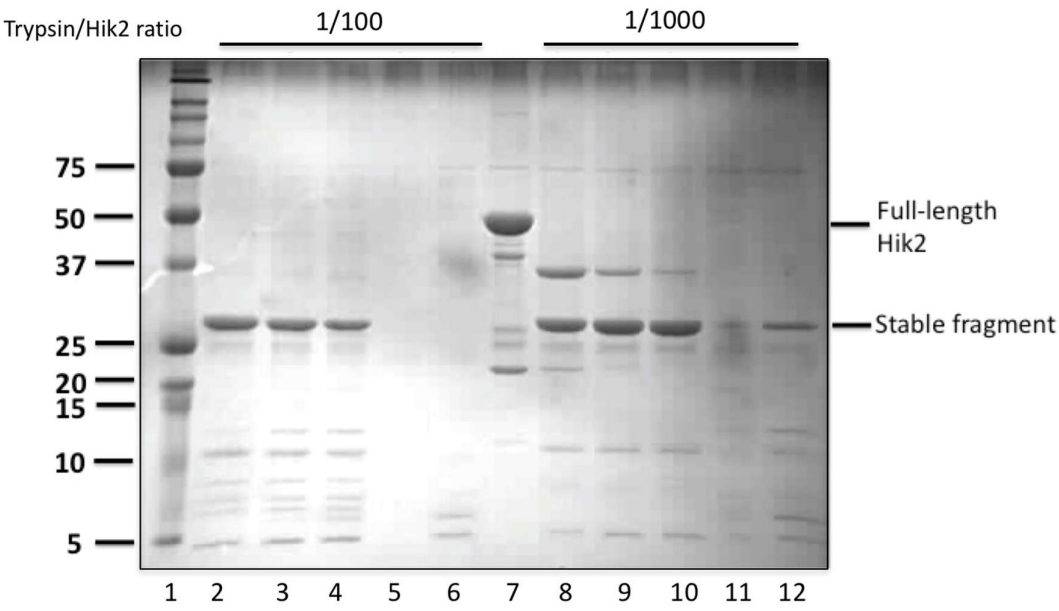


Figure 4.11 SDS-PAGE profile of limited proteolysis for *Synechocystis* Hik2 treated with trypsin. Lane 1, protein molecular mass markers in kDa; lanes 2 to 6 are samples from 20 min, 40 min, 60 min, finally overnight incubation, at 22 ºC, and overnight incubation at 4 ºC, in the presence of 100: 1 protein: trypsin ratio, respectively; lane 7, full-length Hik2 protein sample; lanes 8 to 12 are samples, again, after 20 min, 40 min, 60 min, then overnight incubation at 22 ºC, and overnight incubation at 4 ºC, in the presence of 1000: 1 protein: trypsin ratio respectively. Samples were subjected to a 12% SDS-PAGE separating gel followed by staining with Coomassie blue. The kDa values, for the molecular mass markers used in lane 1, is shown on the left.

These samples were probed by western blotting using an anti-histidine tag antibody and found that the stable fragment did not contain any his-tag (data not shown). Native gels were used to test the solubility of the obtained proteolytic fragment. In Figure 4.12 B lanes 2 to 5, the fragment was revealed to be soluble and not aggregated, and also forming oligomers, given it had been running slightly lower than the untreated Hik2 based on its protein band profile. Since the his-tag was placed at the C-terminal end of the full-length Hik2 protein, and the molecular mass corresponds to its expectation, and

furthermore, it forms oligomers correlating to the presence of the dimerisation-inducing DHp domain, it was inferred that this fragment contained both the GAF and DHp domains.

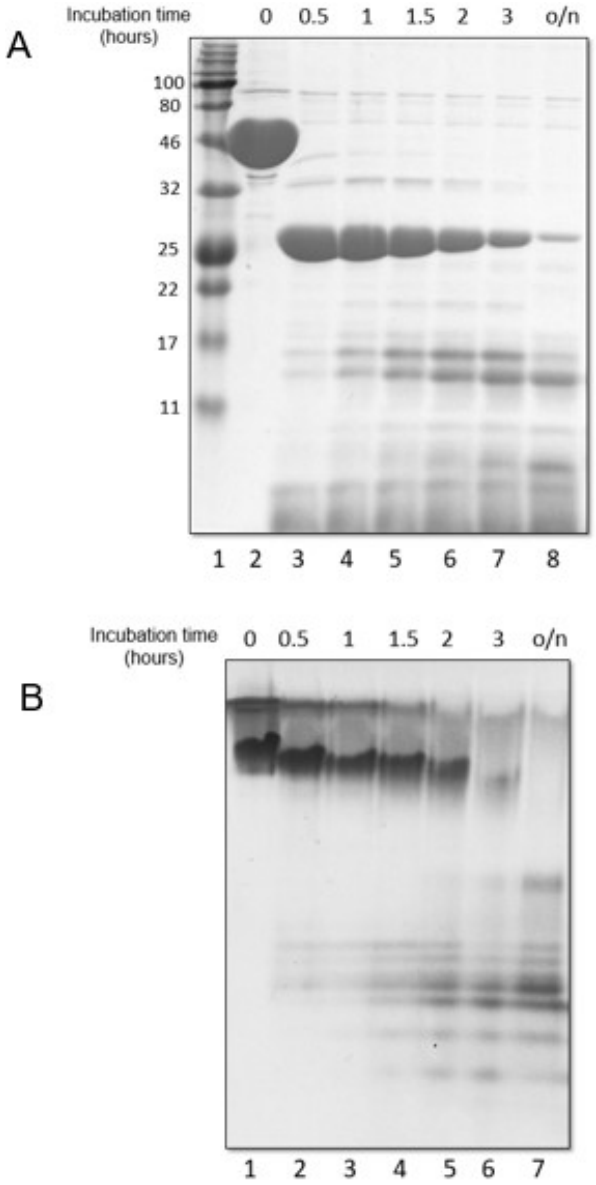


Figure 4.12 SDS-PAGE and native-PAGE profiles of *Synechocystis* Hik2 samples treated with trypsin. A) SDS-PAGE profile of Hik2 after limited proteolysis. The samples were subjected to 12% SDS-PAGE separating gel. The kDa values for the molecular mass markers are shown on the left. B) Native-PAGE profile of Hik2 after limited proteolysis. The trypsin to Hik2 ratio was 1/1000. The incubation time is as indicated, in hours; o/n, overnight.

The stable fragment labelled in Figure 4.12 A, lane 4 was excised from the gel and sent to University of Cambridge to perform N-terminal amino acid sequencing (see section 2.4.3). This fragment turned out to have molecular mass of 27.6 kDa, formed of amino acids 98 to 342, including part of the GAF domain, the whole DHp domain and part of the CA domain (Figure 4.13), indicating that the ^{97}Arg residue in the GAF domain and the ^{342}Lys residue in the CA domain are exposed, making them liable to trypsin digestion.

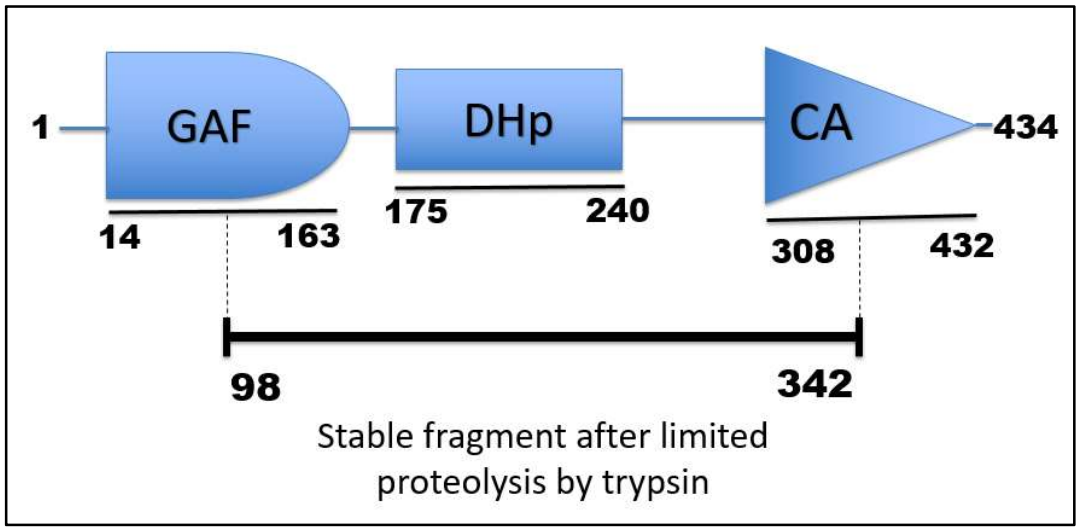


Figure 4.13 Amino acid region of the stable fragment obtained after limited proteolysis of the full-length Hik2 treated by trypsin. The region is from 98 to 342, 245 amino acids in length.

The online server I-TASSER (Zhang 2008) was used to predict the structures of the full-length and subdomains of Hik2, and the digestion sites, labelled in black, are both located on the flexible loop, and should be exposed on the protein surface even when the proteins form a higher order oligomer (Figure 4.14).

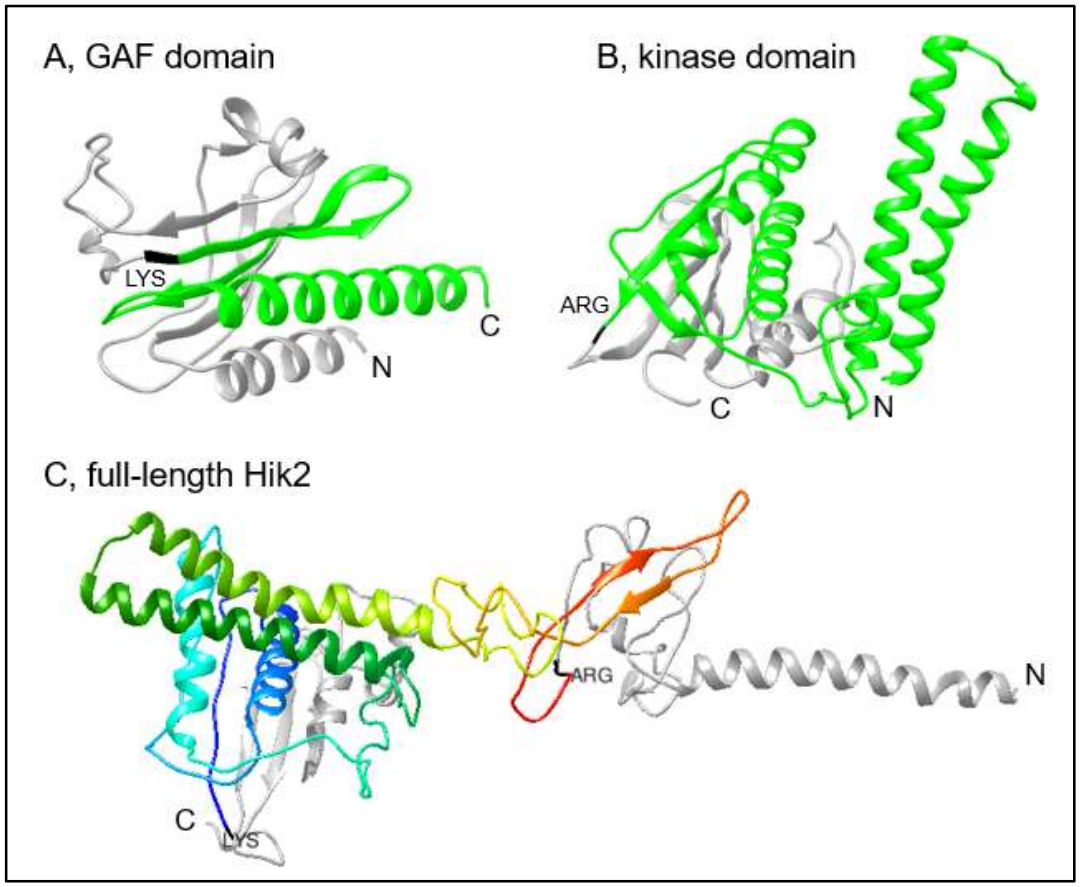


Figure 4.14 The stable fragment region in the predicted Hik2 structure. A) The I-TASSER predicted structure of the Hik2_GAF domain. B) The I-TASSER predicted structure of the Hik2_kinase domain. C) The I-TASSER predicted structure of the full-length Hik2. The stable fragment parts obtained from limited proteolysis are coloured in green in panel A and B, and coloured in rainbow in panel C. The digestion sites ^{97}Arg and ^{342}Lys residues are labelled in black.

4.3 Discussion and future work

4.3.1 The insolubility of the proteins

The solubility of proteins in aqueous solutions is a physicochemical property, and it usually depends on the hydrophilicity of the protein surface. It is also affected by environmental factors, such as the pH, temperature, the ionic strength, and the solvent used. In the 20 common amino acids found in nature, Alanine (Ala), Valine (Val), Leucine (Leu), Isoleucine (Ile), Proline (Pro), Phenylalanine (Phe), Tryptophan (Trp), and Methionine (Met) are hydrophobic, because their side chains have a low propensity to be in contact with water. One exception is that the Cysteine (Cys) residues are usually found to associate with hydrophobic regions of proteins because its side chain can stabilise hydrophobic interactions (Heitmann 1968), even though it is considered to be hydrophilic by itself. The one amino acid without a side chain, Glycine (Gly), is often found within loop or coil regions at the surface of proteins, suggesting it is rather hydrophilic. In a well-folded protein, the hydrophobic residues tend to be located in the protein core, but some also exist in patches on the surface. The distribution of hydrophilic and hydrophobic amino acid residues affects the solubility of proteins, and those with high hydrophobic amino acid content on the surface have low solubility in an aqueous solution.

As is shown in Figure 4.15, there are many hydrophobic amino acid residues in Hik2_GAF and CSK_LBD proteins, rendered in yellow, that are located on the surface, making the proteins rather insoluble in aqueous solution. Moreover, there are several Cys residues in these proteins, thus reducing agent DTT was added to the buffer to avoid disulfide bond formation during the experiments.

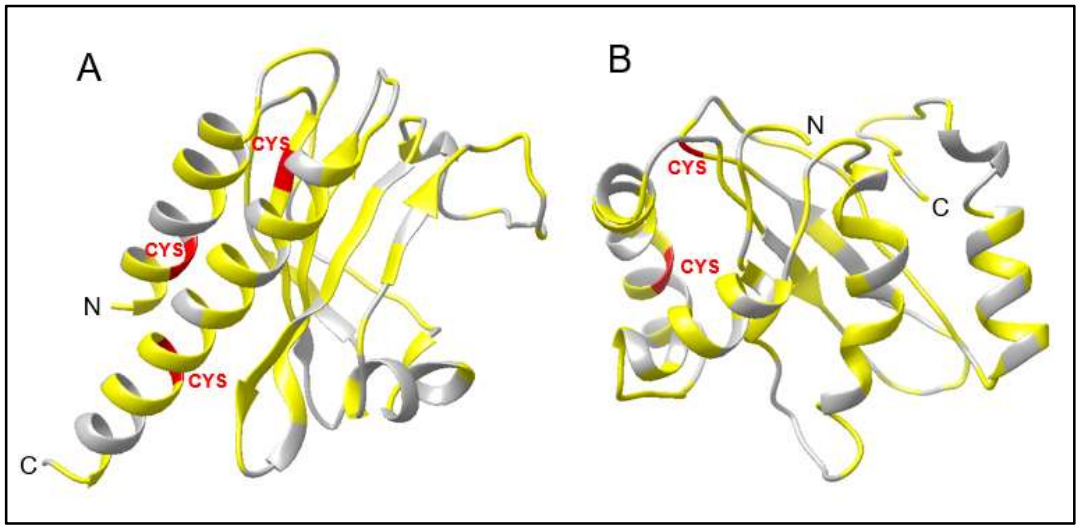


Figure 4.15 The predicted structures of Hik2_GAF and CSK_LBD, showing hydrophobic amino acid residues. A) The Hik2_GAF structure predicted by I-TASSER online server. B) The CSK_LBD structure predicted by I-TASSER online server. The hydrophobic amino acid residues are rendered in yellow, and Cys residues labelled in red.

In order to obtain biologically active and soluble protein, several strategies were tried to improve the solubility of the expressed protein in this project, including the denaturation and refolding of inclusion bodies *in vitro*, and the usage of MBP and SUMO fusion tags. However, the yield of the refolded and cleaved target protein remained stubbornly quite low.

In the future, more crystallisation screening trials should be performed for the MBP-tagged sensor domain proteins, to get the X-ray diffraction of the fusion proteins. Otherwise, since the protein that contains more polar residues is known to have better solubility, the amino acid sequences of Hik2 homologues from different cyanobacterial species could be analysed, and try to express and purify the sensor domains with more hydrophilic amino acid residues.

4.3.2 The signals that Hik2 and CSK sense

The Hik2 homologue in higher plants CSK has been shown to be able to form a quinone adduct, and bind the PQ analogue DBMIB (dibromothymoquinone) with a Kd value (equilibrium dissociation constant) comparable to other quinone binding proteins (Puthiyaveetil et al. 2013, Ibrahim et al. 2016), thus it may act as a redox sensor. It has been reported that a GAF-containing histidine kinase, a bacteriophytochrome-related protein in the photosynthetic bacterium *Rhodopseudomonas palustris*, *Rp*BphP4, acts as a light sensor or redox sensor in different strains, regulating the synthesis of light harvesting complexes (Vuillet et al. 2007). Interestingly, it is suggested that the redox sensing is mediated by the reversible redox-dependent formation of aggregation via intermolecular disulfide bonds of two conserved Cys residues which locate in the PHY domain (a subdomain identified in phytochromes) and CA domains, but not by the GAF sensor domain. The oligomerisation effect of the disulfide bonds in *Rp*BphP4 is correlated to the dimerisation function of the DHp domain in other histidine kinases (Vuillet et al. 2007). There is a Cys residue at position 573 in the CA domain of CSK, which may be responsible to the redox sensing of CSK. Meanwhile, several redox agents, such as potassium ferricyanide, benzoquinone, hydroquinone and DTT, were tested by Dr. Ibrahim, and Hik2 did not respond to them *in vitro* (personal communication, Ibrahim I., 2016), which may be because there is no Cys residue in the CA domain of Hik2 for this redox sensing activity.

As introduced in Chapter 1 section 1.2.1, there are two types of Hik2 protein, and the majority belongs to Class I, those containing both sensor domain and kinase domain, can be found in chloroplasts and most cyanobacteria. The other type, Class II Hik2, consists of only the kinase domain, but no sensor domain, and is found in three marine

cyanobacteria species, the *Gloeobacter violaceus* PCC 7421, *Synechococcus* sp. JA-2-3B'a (2-13), and *Synechococcus* sp. JA-3-3Ab, which diverged very early from other species (Ibrahim et al, 2017). In this thesis, as discussed in Chapter 3, there is possibility that the DHp domain in Hik2 kinase domain can sense Na^+, thus it may be sufficient for the Class II Hik2s in these three species to play their roles with only a kinase domain.

However, the Class I Hik2s that still preserve a sensor domain, which are mostly conserved GAF domains, raises a question: what other signals do they sense? The solved structures of GAF domains indicate that GAF domain can bind cGMP and cAMP for the cyclic nucleotide signalling (Soderling and Beavo 2000), or bind a bilin chromophore for the light absorption (Wagner et al. 2005, Wagner et al. 2007), or bind a heme for the O_2 detection (Roberts et al. 2004), etc. It is possible that the GAF domain in Hik2 also binds small ligands for the regulation of its autophosphorylation activity. It has been reported that the GAF domain of Hik2 senses Cl^- concentration and responds to salinity stress *in vivo*. This was deduced by examining the alkaline phosphatase activity of a reconstructed chimeric protein complex containing both the Hik2 GAF domain and the Hik7 kinase domain (Kotajima et al. 2014). It should be noted that Cl^- plays an important role in photosynthesis, e.g. it was observed from crystal structures that PS II directly binds Cl^- (Kawakami et al. 2009, Umena et al. 2011), and chloride depletion suppresses PS II to transfer protons to the lumen (Rivalta et al. 2011). Furthermore, Cl^- may serve as a counterion of Na^+ and K^+, regulating the osmolarity of the native environment (Hagemann 2011). This also corresponds to the prediction of Hik2 functioning in salt/osmolarity sensing and in a signalling pathway in photosynthesis (Ibrahim et al. 2016). However, the Cl^- binding site has not been determined yet for the GAF domain of Hik2, since all tested mutants of the seven

conserved amino acid residues in the GAF domain had lost Cl^- sensing activity (Kotajima et al. 2014). This might eventually be elucidated by X-ray diffraction studies of the GAF domain crystal, with bound Cl^- ions, in the future.

4.4 Conclusions

The sensor domain of a histidine kinase usually perceives environmental signals, allowing them to be transduced to a cognate response regulator, and subsequently introducing a proper response to this stimulation. In this chapter, the sensor domains of *Synechocystis* Hik2 and *Arabidopsis* CSK were analysed, overexpressed, and *in vitro* denaturation and refolding, or ligation with MBP or SUMO tags, were performed, in order to purify and solubilise them, for the purpose of getting their structural information from X-ray crystallisation or NMR spectroscopy. However, they proved stubbornly, and disappointingly, insoluble. The possible reasons and potential solutions have been discussed. However, 3D models for these two sensor domains were predicted by online modelling servers, and in this context the signal sensing by Hik2 and CSK was discussed. Also, limited proteolysis was performed on *Synechocystis* Hik2, and additional tertiary structural information was obtained in order to aid the overall predicted outcomes.

Chapter 5 Transmission electron microscopy and single particle image processing of full-length Hik2

5.1 Overview

5.1.1 Transmission electron microscopy

Transmission Electron Microscopy (TEM) is a biophysical technique enabling one to observe small objects or sections of objects, typically under 1 μm in size, such as cells, viruses, and protein molecules. Combined with the image processing technique of single particle analysis, it can offer direct visualisation of biological specimen details, and has a significant advantage for those proteins that are found to be difficult to crystallise, in order to gain X-ray diffracted structural information. There are several components that make up the standard TEM design; an electron emission source to generate the electron stream, a pumping system so that the electrons may travel through a vacuum, a specimen-controlling stage and an imaging device (Reimer 2008) (Figure 5.1).

The basic operating principle of TEM is demonstrated in Figure 5.1. After being generated by the electron source, the electron beam goes into the electromagnetic condenser lens system, which only allows the electrons within a small energy range to pass through and excludes high angle electrons, thus the beam becomes tightly focused. Then the electrons are scattered when passing through the specimen, and are again focused by an electromagnetic projection objective lens system, generating a diffraction pattern or an image on a fluorescent screen or a Charge Coupled Device (CCD) camera (Reimer 2008, Mielańczyk 2015).

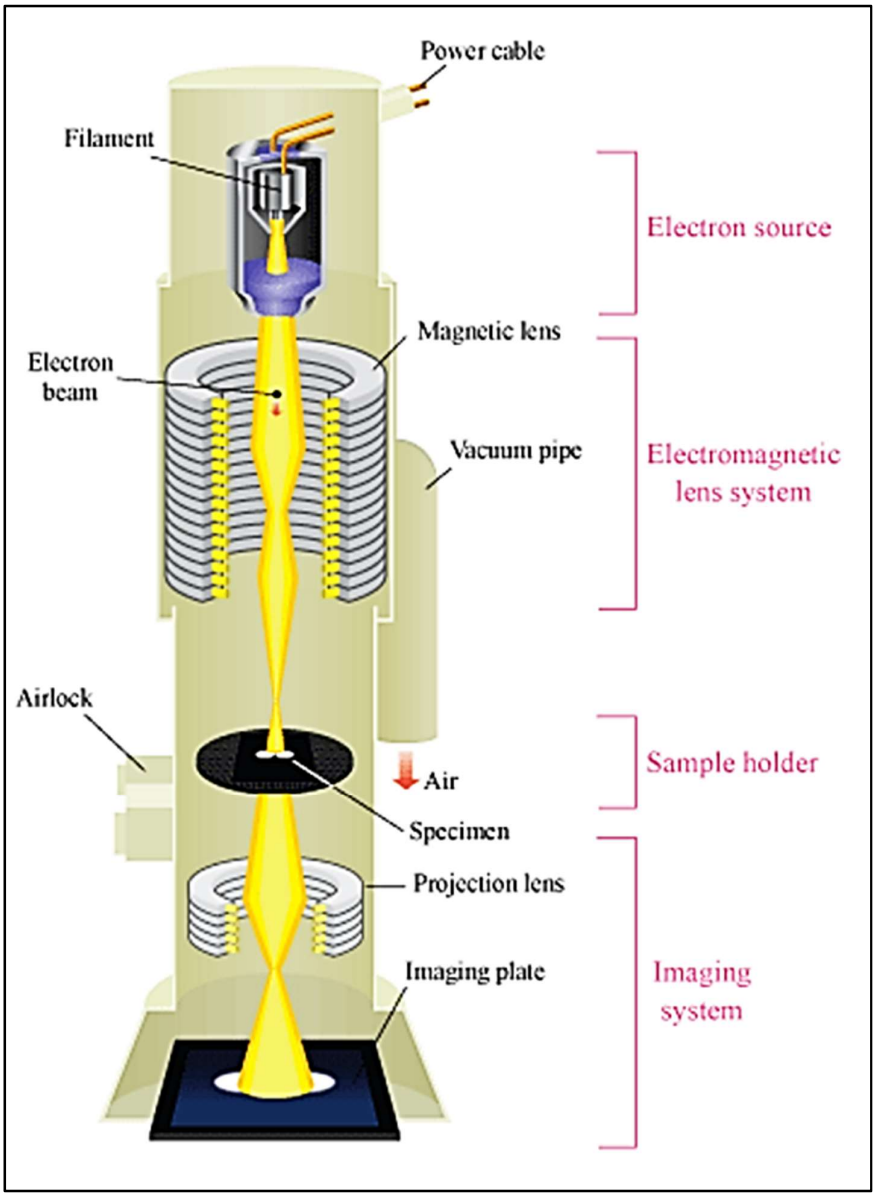


Figure 5.1 The schematic outline of a typical TEM. Figure from (Atomic world, http://www.hk-phy.org/atomic_world/tem/tem02_e.html).

Since the wavelength of electrons is much smaller than that of visible light, a much higher resolution can be obtained through the use of TEM compared to that of optical microscopes, where the physics of image formation in the TEM, when utilizing an electron beam at, for example, an accelerating voltage of 100 kV, is usually in the order

of picometres. At this point it is the technology and development of the magnetic lenses and instrumentation involved which has limited the best TEM investigations to be in the region of 0.5 to 5 Å (Kisielowski et al. 2008). For example, the structure of human 80S ribosome has been resolved at a resolution of 3.6 Å by cryo-EM operating at 300 kV acceleration voltage at a magnification of 79,000 × (Myasnikov et al. 2016); the structure of a protein complex, the crRNA-guided surveillance complex (Csy complex) from *Pseudomonas aeruginosa* in association with two distinct anti-CRISPRs from *Pseudomonas* phage, was determined at 3.4 Å by cryo-EM operating at a magnification of 79,000 × (Chowdhury et al. 2017).

5.1.2 TEM imaging and processing

To gain better, more detailed resolution, the electron micrographs with minimal astigmatism and drift are usually selected by the inspection of optical diffraction, for the subsequent image processing (Reimer 2008, Carazo et al. 2015). Under the electron beam, the amorphous carbon film of the grid produces a nearly white spatial frequency spectrum, but not all spatial frequencies are transmitted equally well, and the missing frequencies appear as Thon rings in the optical diffraction pattern or the Fourier transform (Reimer 2008). As shown in Figure 5.2 A, the power spectrum of a typical bright field TEM image of amorphous carbon film presents concentric Thon rings, and the corresponding radial intensity of the power spectra is shown in Figure 5.2 B. The grey Thon rings indicate weak signal, caused by the minima in the Contrast Transfer Function (CTF, to be discussed below), while the white rings correspond to the contrast transfer maxima. The Thon rings with rotational symmetry show low astigmatism, while the stretched rings indicate the image is astigmatic. The beam drift during exposure makes the Thon rings blurring. Therefore, electron micrographs should be

routinely inspected by computer and software imaging suites, or physically by optical diffraction, before being taken forward for subsequent averaging analysis. Also, the spacing of the Thon rings can act as a feedback aid to estimate the defocus value when taking later micrographs.

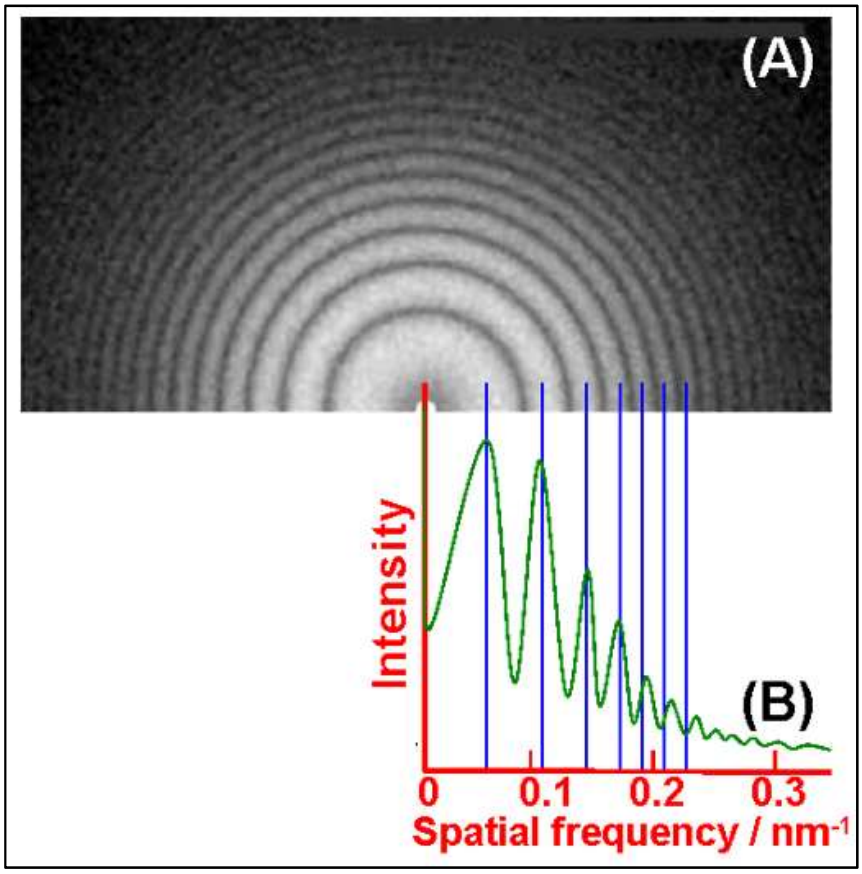


Figure 5.2 The Thon rings and the corresponding radial intensity. A) The power spectrum of a typical bright-field image of amorphous carbon film presenting concentric Thon rings. B) Radial intensity of the power spectra. Figure from (Liao 2007).

When the electrons pass though the objective aperture and the specimen, they are elastically scattered, and interfere with the unscattered transmitted wave, producing phase contrast. There are also electrons being inelastically scattered and losing energy, or falling outside the objective aperture by high angle elastic scattering, which causes

amplitude contrast. Both phase and amplitude contrast provide the contrast of a TEM image (Amos et al. 1982).

The CTF mathematically describes the aberrations of TEM modification to the images caused by the non-equally transmitted spatial frequencies as mentioned before (Reimer 2008). High-resolution information can be obtained from the high spatial frequencies, and low spatial frequencies represent coarse detail (Figure 5.2 B). The CTF correction is applied to retrieve the undistorted object from TEM images by using software packages such as CTFFIND (Rohou and Grigorieff 2015), RELION (Scheres 2012), EMAN2 (Ludtke et al. 1999), Imagic-5 (van Heel et al. 1996) or SPIDER (Frank et al. 1996).

5.1.3 Single particle image averaging/analysis

Single particle image averaging, also termed single particle analysis, is a computerised image-processing/averaging technique that has been developed over the past ~ 40 years which may be applied to analyse images from TEM. It is deployed to improve and extend the information obtained from TEM images of particulate samples, typically proteins or other large biological entities (Ruprecht and Nield, 2001), and in the past few years has become a mature technique, even revealing atomic details in large, macromolecular, example systems that have eluded ready crystallisation, precluding diffraction studies, of this level of resolution quality may be found for the *E. coli* ribosome at 2.9 Å overall resolution (Fischer et al. 2015) and the human ribosome at 3.6 Å (Myasnikov et al. 2016), the human 20S proteasome at 1.8 Å resolution (Schrader et al. 2016), and the *Pseudomonas* crRNA-guided surveillance complex at 3.4 Å (Chowdhury et al. 2017).

To readily obtain high contrast TEM images, protein samples should be negatively stained, for example, by using 1 to 2% (w/v) Uranyl Acetate (UA) or Phosphotungstic Acid (PTA), and then imaged using a TEM at 80 kV and a magnification between 50,000 × to 80,000 × (or up to 120,000 ×, size of molecule and resolution expectations considered). As many micrographs for each sample as possible, displaying no discernible drift or astigmatism, are then digitised by a CCD, CMOS (complementary metal-oxide semiconductor), or direct electron detection camera systems. This should result in a sampling frequency of the order of 1 to 10 Å per pixel, at the scale of the specimen, which has enough contrast to be applied subsequently to a range of software environments developed specifically for single particle analysis, such as the Imagic-5 software (van Heel et al. 1996, Ruprecht and Nield 2001), SPIDER (Frank et al. 1996), RELION (Scheres 2012), Xmipp (Sorzano et al. 2004), and more software can be found on website http://3dem.ucsd.edu/software.shtm. The CTF modulates the amplitudes and phases of the electron diffraction pattern formed in the back focal plane of the objective lens, but no correction is required if the first minimum of the power spectrum for each micrograph is likely to be in the 17 Å to 21 Å range. Relatively large heavy metal molecules are used in high contrast negative stains, and these are unable to penetrate the protein structure, thus the resolution for such studies is limited to ~ 12 Å to 15 Å (Kiselev et al. 1990).

To aim for 12 to 15 Å in resolution, a data set of many thousands of particles should be built, ideally > 10,000 individual molecular protein complex images, by interactively selecting all possible single particles observed. In the later stages, where the views of the protein become understandable, this selection process may be anticipated, e.g. by using the ‘boxer’ program of EMAN2 (Ludtke 2010). The data set built is then analysed

by reference-free alignment coupled to multivariate statistical analysis (van Heel et al. 1996). If the sample is homogeneous, a number of subpopulations within each data set will display at least some top, side, or intermediate orientations. Each of these subpopulations can be treated *de novo* and the families, or classification, of these views, are then iteratively refined to gain two-dimensional (2D) class averages with improved signal-to-noise ratio (Harauz et al. 1987). 2D reprojections taken from the best three-dimensional (3D) model obtained are used to identify additional atypical views and thus further refine the subpopulation class averages.

For 3D reconstruction, taking a broad range of class averages from each 2D subpopulation will ensure the largest range of relative orientations, so that Eulerian angles may be assigned *a priori* by angular reconstitution (Van Heel 1987) and iterative refinement can lead to the resolution of one, or possibly several, final 3D map(s), composed of tens or hundreds of class averages from each subpopulation data set. These would then be estimated by Fourier Shell Correlation (FSC) to gain an approximate resolution value (van Heel et al. 2000, Liao and Frank 2010).

Modelling studies may be performed within modelling packages such as PyMOL (The PyMOL Molecular Graphics System, Schrödinger, LLC) or UCSF Chimera (Pettersen et al. 2004). The key steps of single particle analysis, image processing and 3D reconstruction have been demonstrated elegantly in Figure 5.3 (Ruprecht and Nield 2001).

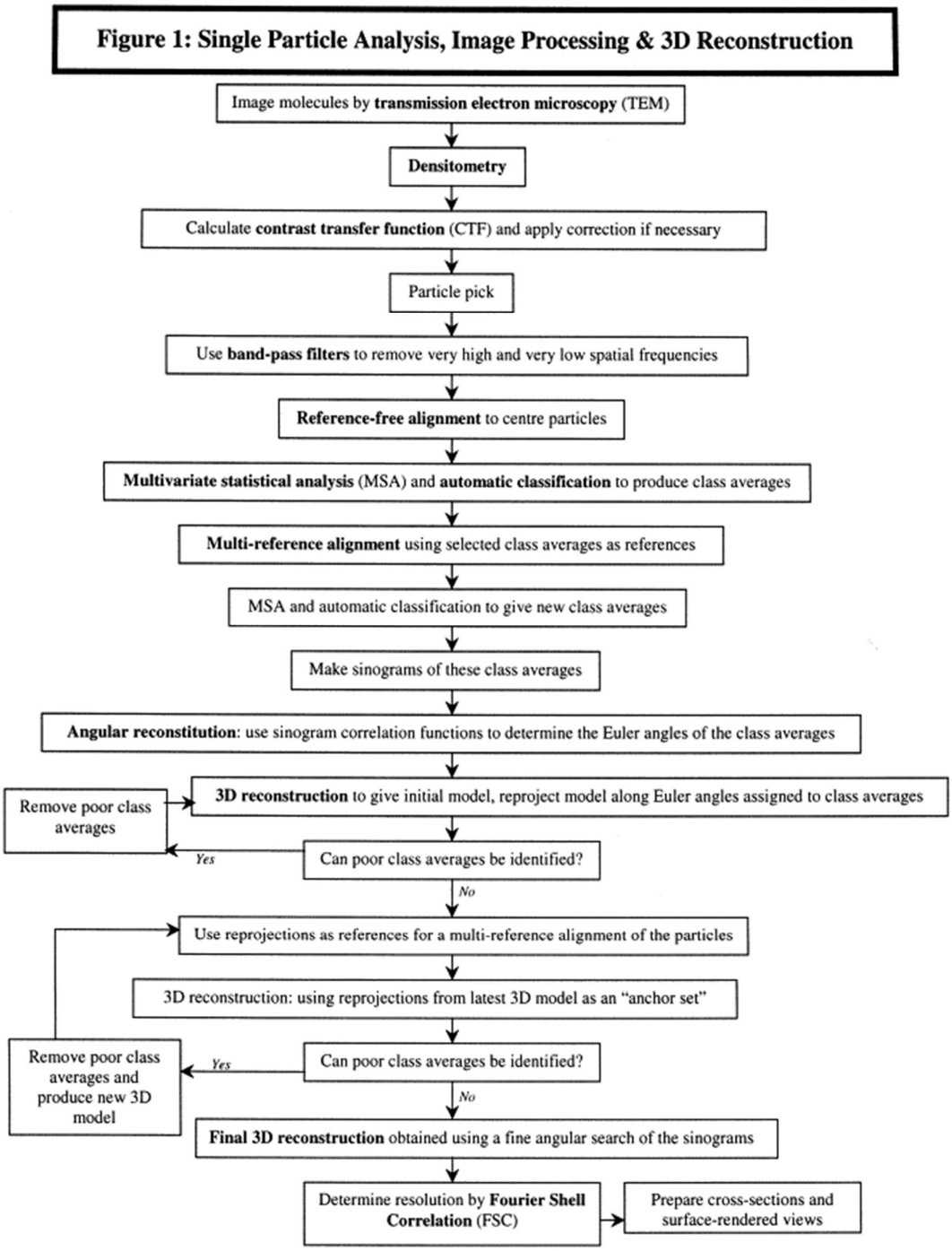


Figure 5.3 The key steps of single particle analysis, image processing and 3D reconstruction. Figure from (Ruprecht and Nield 2001).

In this project, based on the results discussed in Chapter 3, an equilibrium in Hik2 solution between monomer, tetramer, and hexamer was anticipated. Since direct visualisation of complexes may be obtained by TEM, the full-length *Synechocystis* Hik2 protein sample was applied to TEM and the images processed by single particle analysis, in order to study the oligomeric state and structural information for Hik2. It was expected that protein particles with different sizes, from the size of monomer with the molecular mass of around 50 kDa to hexamer around 300 kDa, would be observed.

5.2 Results

5.2.1 TEM imaging and single particle analysis for the Hik2 protein from *Synechocystis* sp. PCC 6803

Hik2 protein from *Synechocystis* sp. PCC 6803 was expressed and purified using the method described in Chapter 2 section 2.2, then applied to size exclusion chromatography (elution buffer was 20 mM Tris-HCl (pH 7.5), or 20 mM Tris-HCl, 10 mM NaCl (pH 7.5)), where the dominant form of the protein was found to be hexameric with a molecular mass of approximately 290 kDa. Initial training was performed with Dr. Jon Nield and Dr. Kristina Zubow, then a number of independent samples from Hik2 *Synechocystis* sp. PCC 6803 were observed under the electron beam using a TEM manufactured by JEOL, Japan (originally Japan Electron Optics Laboratory Company, Tokyo, Japan), model 1230, with an accelerating voltage of 80 kV and imaged at 80,000 × magnification, as housed within the NanoVision Centre of Queen Mary, University of London.

After these training, 40 individual TEM sessions were performed alone during the past three years. In the preliminary sessions, interesting characteristic views of Hik2 in different concentrations were observed, as shown in Figure 5.4. Interestingly, when diluted in H_2O (Figure 5.4 A), the protein appeared better stained in the central water drop-like area, with higher contrast, while in Figure 5.4 B, many black particles with different sizes and shapes can be seen, but no protein. One possible reason may be that the stain was not filtered properly, and the black particles are PTA stain crystals, while proteins had become buried in the grey background with low contrast compared to the PTA crystals.

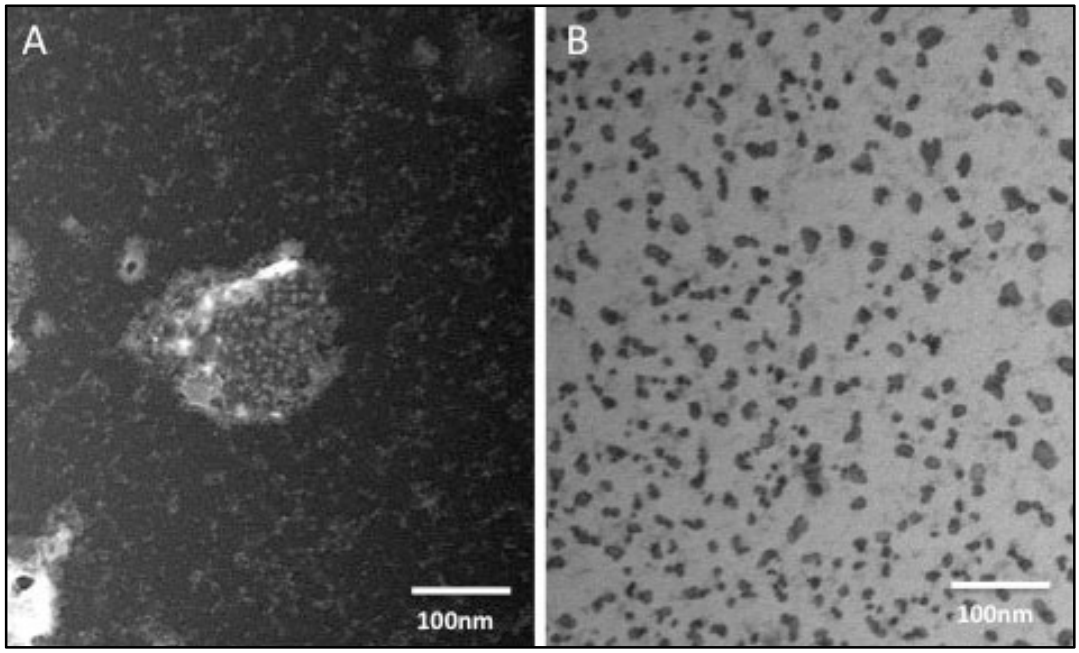


Figure 5.4 TEM images of Hik2 samples embedded in 2% PTA negative stain. A) Hik2 protein diluted in H_2O at 2.5 μg/ml (obtained from grid A3 on 21/01/2012). B) Hik2 protein in buffer with 20 mM Tris-HCl (pH 7.5) at 2.5 μg/ml (obtained from grid A9 on 08/02/2013), however, given the positive apparent, this was attributed to crystals of PTA, not protein. Scale bar, 100 nm, as shown. Micrographs were taken and supplied by Dr. Zubow.

After the initial 10 sessions, 30 promising images were obtained, where Dr. Nield processed a first batch of images, chosen for quality on the basis of minimal observed drift and astigmatism, and single particle analysis was performed together as part of the overall training. Figure 5.5 shows a typical region within one of those micrographs and nine exemplar 2D class averages, taken from a total set of 900 class averages, having first been classified into families of like size and orientation by multivariate statistical analysis. The total data set was 9,132 particles from these sessions.

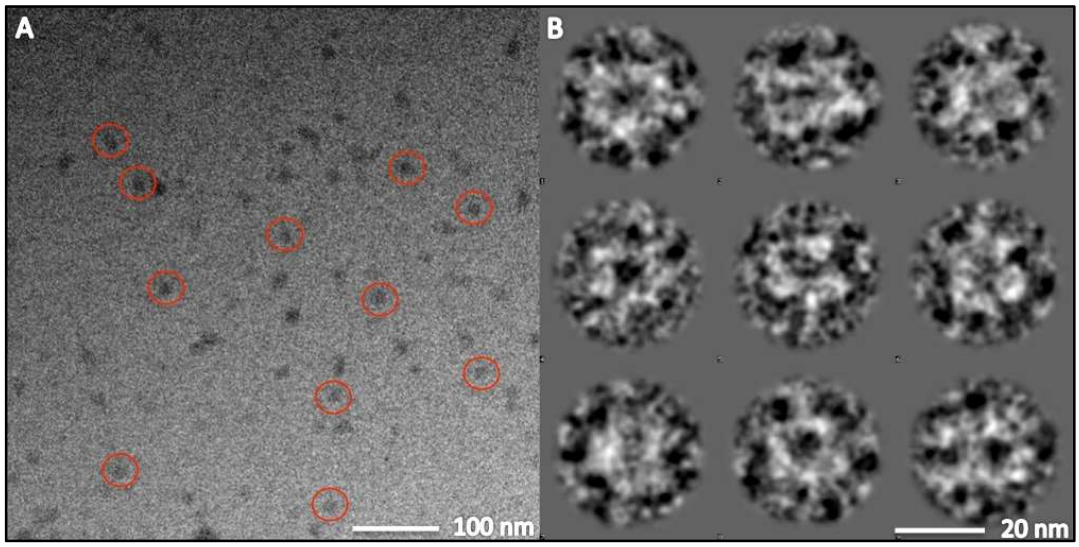


Figure 5.5 TEM image and preliminary single particle analysis of Hik2 protein. A) Typical region of a micrograph of negatively stained Hik2 protein, using PTA stain, obtained from grid E9 on 04/07/2013. Hik2 protein is in buffer with 20 mM Tris-HCl (pH 7.5) at 2.5 μg/ml. Hik2 particles are circled. B) A selection of nine characteristic 2D views of Hik2. Protein is black in A, and inverted to white during processing in B; scale bars, respectively, 100 nm and 20 nm, as shown.

5.2.2 TEM-derived 3D maps of Hik2 complexes and comparisons with 3D homology modelling by the I-TASSER algorithm

The I-TASSER (Iterative Threading ASSEmbly Refinement) server is an online service developed (Zhang 2008) since 2008 for 3D protein structure and function studies. 3D models are built based upon multiple-threading alignments on a so-termed LOMETS (Local Meta-Threading-Server) (Wu and Zhang 2007) and iteratively refined template fragment assembly simulations. The *Synechocystis* Hik2 sequence (GenBank: BAK49478.1) was used and uploaded to the I-TASSER server; five models were reported at the end of ~ 35 h computer job, in which the model shown in Figure 5.6 (red) was the closest match according to core correlation criteria (Pettersen et al., 2004).

The most highly resolved TEM-derived 3D map was calculated by Dr. Nield, based on the data obtained from 3 sessions, each in terms using an independent sample. The green mesh in Figure 5.6 represents the current electron defined molecular density envelope of Hik2, derived from the TEM imaged data and calculated by angular reconstitution using the Imagic-5 software environment (van Heel et al. 1996, Ruprecht and Nield 2001). The 3D map was observed at the same scale with the I-TASSER predicted PDB file for *Synechocystis* Hik2, as shown in Figure 5.6. According to the relative dimensions between the two, it would appear that the TEM 3D model can accommodate 4 to 6 Hik2 monomers. In scale, the overall dimensions of the TEM map are ~ 20 nm × 20 nm × 20 nm, which leads to the assumption that an oligomeric structure may be present, consistent with previous results (such as the ones obtained from size exclusion chromatograph and dynamic light scattering, described in Chapter 3), however, the symmetry of this arrangement proved unclear. The estimated resolution of the TEM model was calculated to be 30 Å by FSC (van Heel et al. 2000).

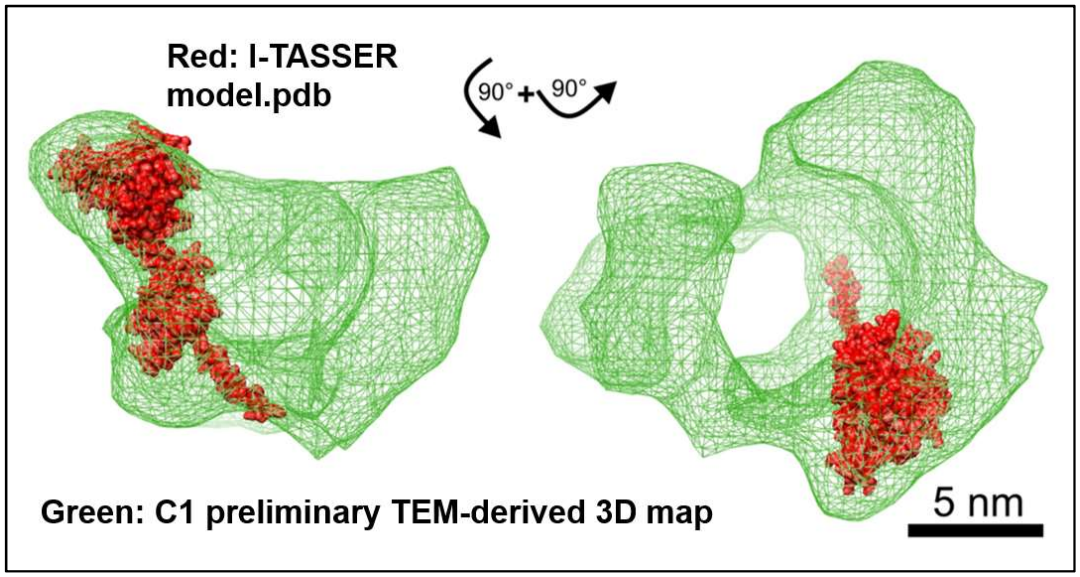


Figure 5.6 TEM-derived preliminary Hik2 3D map incorporating the predicted Hik2 monomer structure. Overall scaling comparisons between the homology model and the TEM 3D map, derived from angular reconstitution of 2D averages (Figure 5.5 B) are shown. The density obtained may accommodate 4 to 6 monomers. Green, C1-symmetric (asymmetry) TEM-derived 3D map; red, Hik2 homology model, depicted as surface rendered atom spheres, as reported by I-TASSER. Scale bar, 5 nm, as shown. Figure supplied by Dr. Nield, and displayed at the 16th International Congress on Photosynthesis (poster), with subsequent student fellowship awarded by the organisers.

Gaining an indication (Figure 5.6) that the Hik2 samples what observation were indeed oligomeric about 20 nm^3 in volume, perhaps containing as many as 4 or 6 monomeric complexes, motivated continued trials of TEM imaging, in order to more fully understand their oligomeric states.

5.2.3 TEM imaging of the Hik2 protein and single particle analysis on negatively stained Hik2 oligomers

After several trials of optimising staining conditions, an improved negatively stained grid, in this case labelled M8, provided for 40 images, and 23 images were chosen on the basis of their consistent underfocus close to the TEM-CCD camera systems quality limit (~ 15 Å), lack of drift and low astigmatism; this led to further single particle analyses building upon that of the previous section. Figure 5.7 shows an original micrograph (a square of 2672 × 2672 pixels) of the underfocus obtained. The sampling frequency of the Olympus (Olympus, Japan, Ltd.) Morada system's output was 5.962 Å/pixel when at 80,000 × magnification (see Chapter 2 section 2.7).

The best possible resolution therefore, taking into account of the Nyquist–Shannon theorem (Shannon 1949), will be no better than double the sampling frequency, i.e. 11.924 Å. Unfortunately, the remaining 17 micrographs out of the 40 were not at this 'ideal underfocus'; as well as the limitations of using negative stain and its lack of penetrating ability into the protein structure (Kiselev et al. 1990), where it is not expected to gain details better than 15 Å resolution. Usually an underfocus, where the first minima of the CTF may be calculated in real time by the CCD system, through recording by a live camera feed displaying the Fast Fourier Transform (FFT), and the user can freely choose to focus by turning the focus knob of the TEM console. This avoids the internal detail within the particles being lost through the merging of images that have inherently worse focus. Also, it is quite likely that the protein has degraded in the stain buffer. This may give images with a high level of broken protein background. Thus a resolution at this ~ 15 Å limit was unexpected, however, a resolution of 30 Å was calculated by 2D Fourier Ring Correlation (FRC) methods (Saxton and Baumeister

1982, van Heel 1982), and the size and pattern of the particles were reproducible in all following sessions.

As can be seen in Figure 5.7, the single particles present different sizes and shapes, which is highly indicative of the protein being heterogeneous in the eluted samples (see Chapter 3 section 3.2).

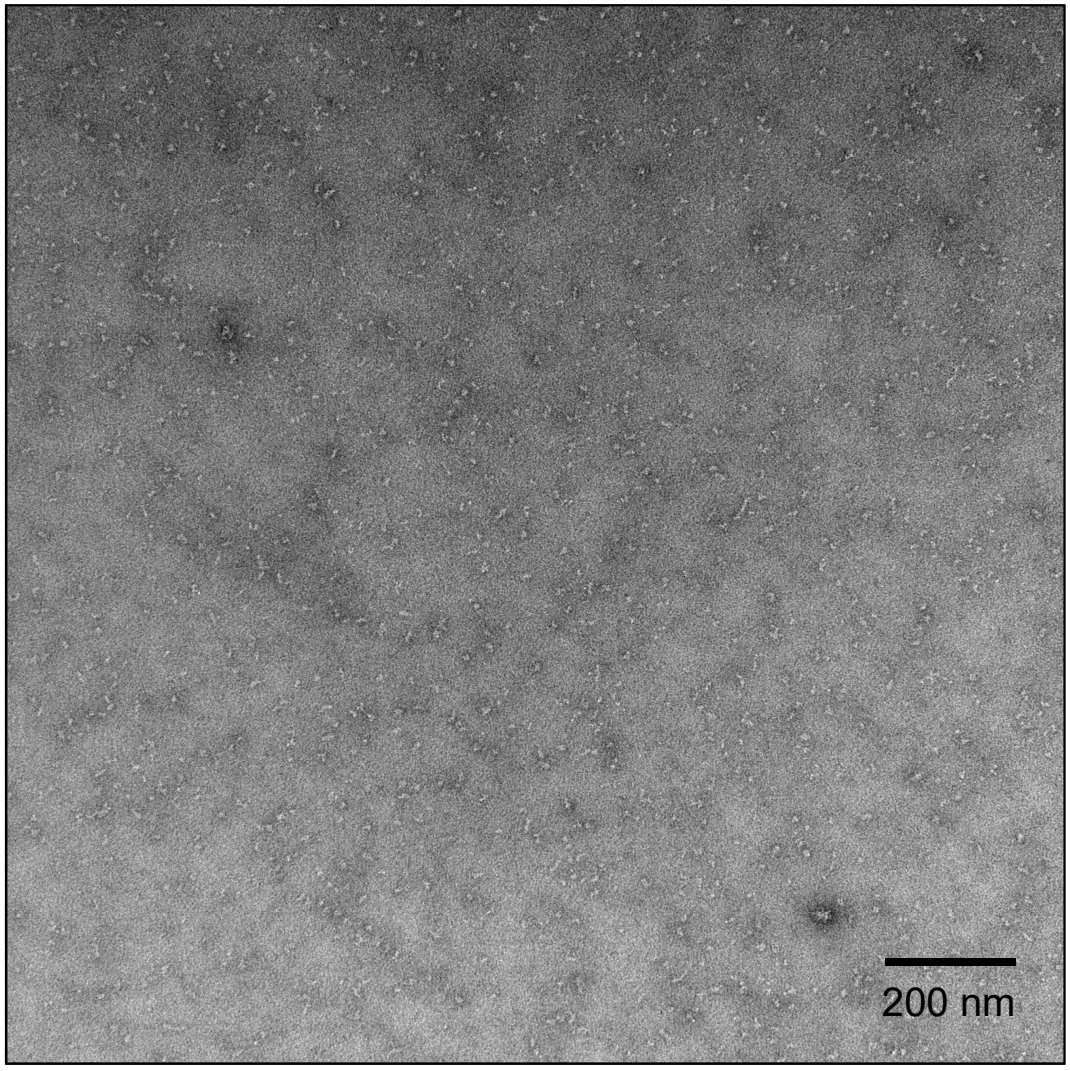


Figure 5.7 A typical TEM micrograph (CCD image) of *Synechocystis* Hik2. The protein was buffered with 20 mM Tris-HCl (pH 7.5) at 5 μg/ml, and embedded in 2% PTA, negative stain (image obtained from grid M8, 28/01/2016).

During these sessions, different concentrations of Hik2 were examined, to find an optimised protein concentration for single particle analysis. As shown in Figure 5.8, compared with the well separated sample at 2.5 μg/ml (Figure 5.8 A), the particles at a concentration of 10 μg/ml appear to associate together, and it was difficult to interactively box them out (Figure 5.8 B**)**. 10 mM NaCl was added to the buffer to maintain ionic strength.

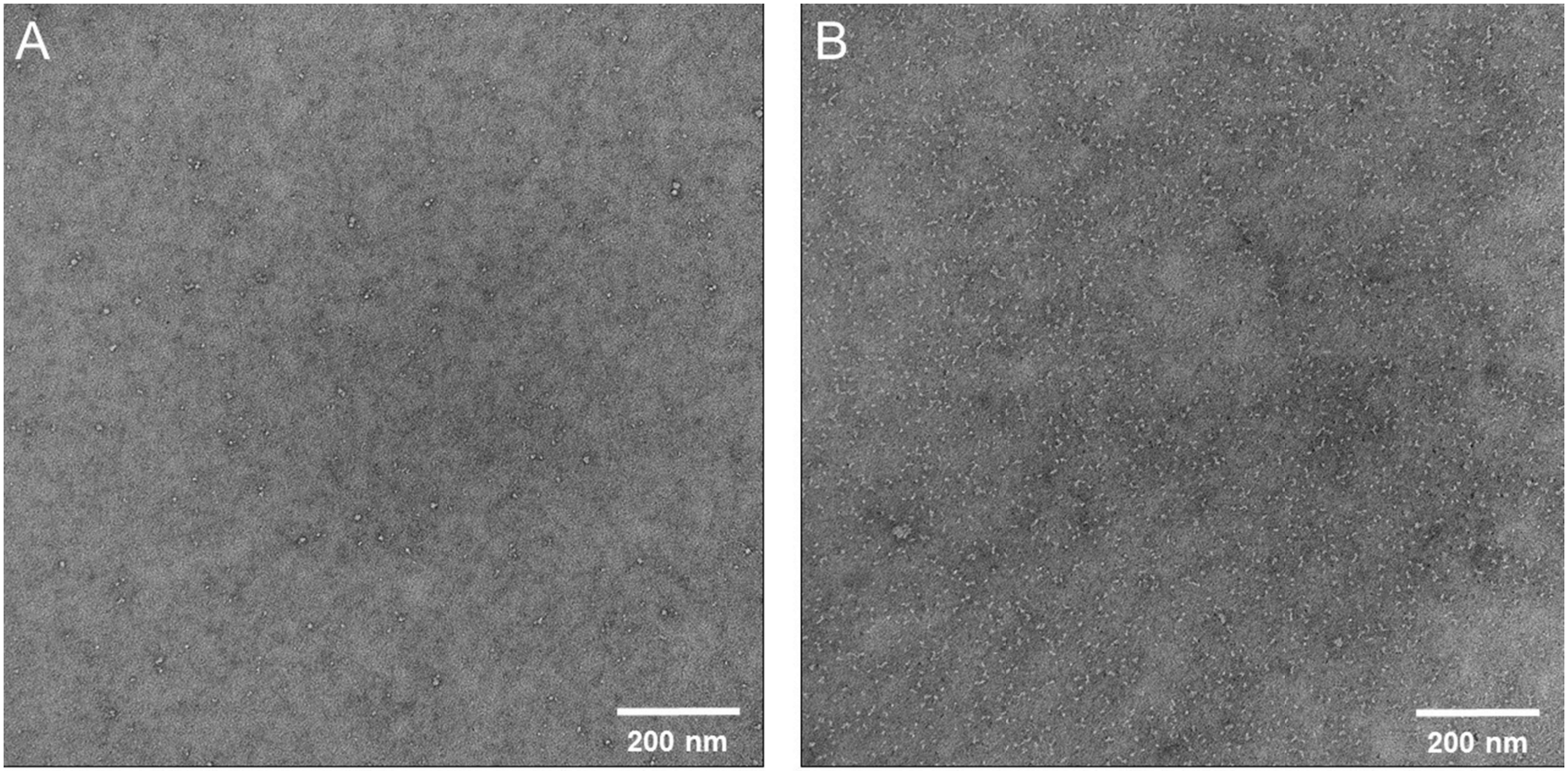


Figure 5.8 TEM images of Hik2 samples at two different concentrations.

The protein was diluted to (A) 2.5 μg/ml and (B) 10 μg/ml. Both proteins are in buffer with 20 mM Tris-HCl, 10 mM NaCl (pH 7.5), embedded in 2% PTA negatively stain (obtained from grid Q2 (A) and grid Q4 (B) on 15/08/2016). Radiation damage (tiny black dots) can be seen on the images.

From the 23 selected images obtained from session 20, a data set of 4,060 particles were selected, and averaged into 250 class sums/characteristic views after 4 rounds classification. Figure 5.9 shows 49 out of the 250 class sums. Some stacking can be seen e.g. 139, 212, and possible their rectangular side views e.g. 112.

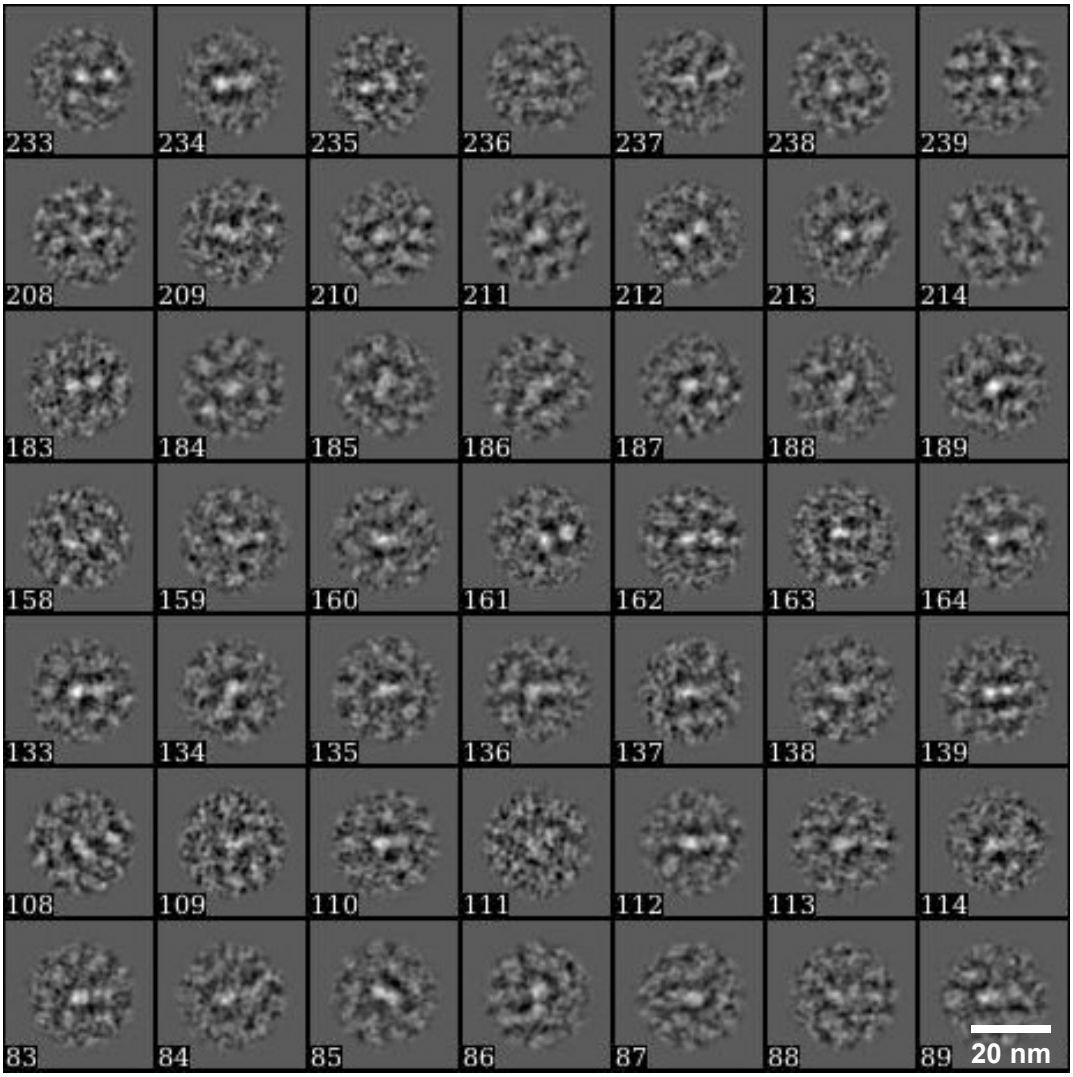


Figure 5.9 Part of 250 classums after 4 rounds of classification. The side of each box is 64 pixels, equivalent to 382 Å. Figure supplied by Dr. Nield.

Afterwards, different negative stain methods were tried with the Hik2 samples, to improve the contrast of the images. Another batch of samples negatively stained by 2%

UA gave better contrast, as can be seen in an original TEM micrograph of Figure 5.10. The colour of the single particles of proteins is strikingly different from the dark grey background. Once again, a variety of different sizes and orientations can be observed, each sample being structural heterogeneous in a similar fashion, which is consistent with previous images and the biochemical results (see Chapter 3 section 3.2).

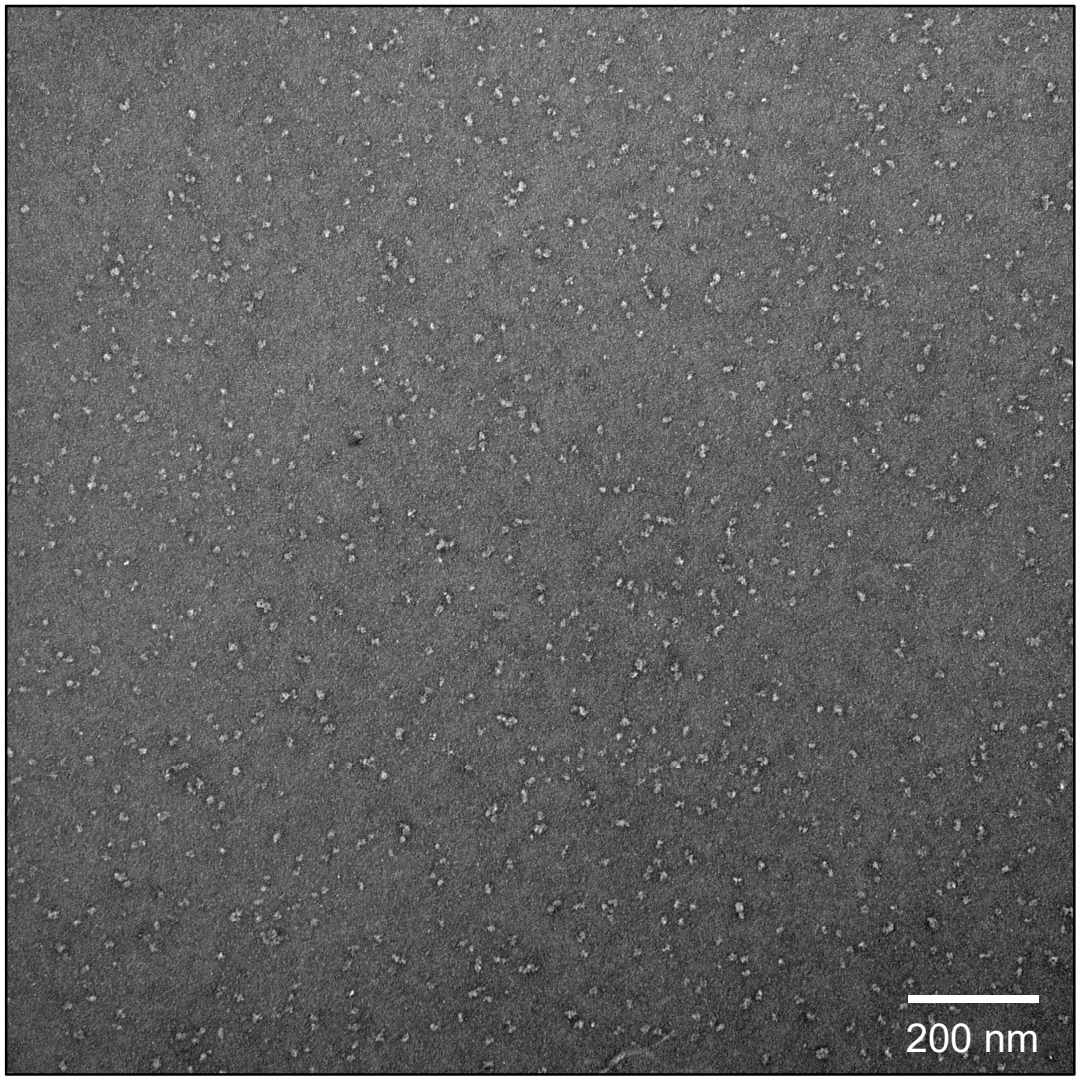


Figure 5.10 A typical TEM micrograph of *Synechocystis* Hik2 with higher contrast. The protein was buffered with 20 mM Tris-HCl, 10 mM NaCl (pH 7.5) at 5 μg/ml, and embedded in 2% Uranyl Acetate (UA), negative stain (image obtained from grid S1, 24/11/2016). Scale bar represents 200 nm.

Together with other images obtained from 4 more independent sample preparations, negative stained by 2% UA, 41 micrographs (of a total of 201 micrographs) were chosen for further single particle analysis, again with the guidance of Dr. Nield. Single particle complexes were floated out into boxes of 64 × 64 pixels in size. No correction for the CTF was applied (see section 5.1.3) and the final class averages presented were low band-pass filtered to 20 Å resolution.

Initial single particle images were selected using the 'boxer' module of EMAN2 (Ludtke et al. 1999). Several distinguishable single particles were chosen and boxed manually as suitable templates, and then the boxing algorithm of EMAN2 was directed to automatically pick all possible single particles in one micrograph, without band-passing or normalising. A typical working window of single particle boxing using EMAN2 'boxer' module is present in Figure 5.11. Window A demonstrates that after manually select several single particles with the black boxes, grey boxes would automatically appear and box similar particles. All boxed particles are magnified in window C, and can be deleted if not satisfied, such as the ones overlapped with others, or those at the border of the micrograph.

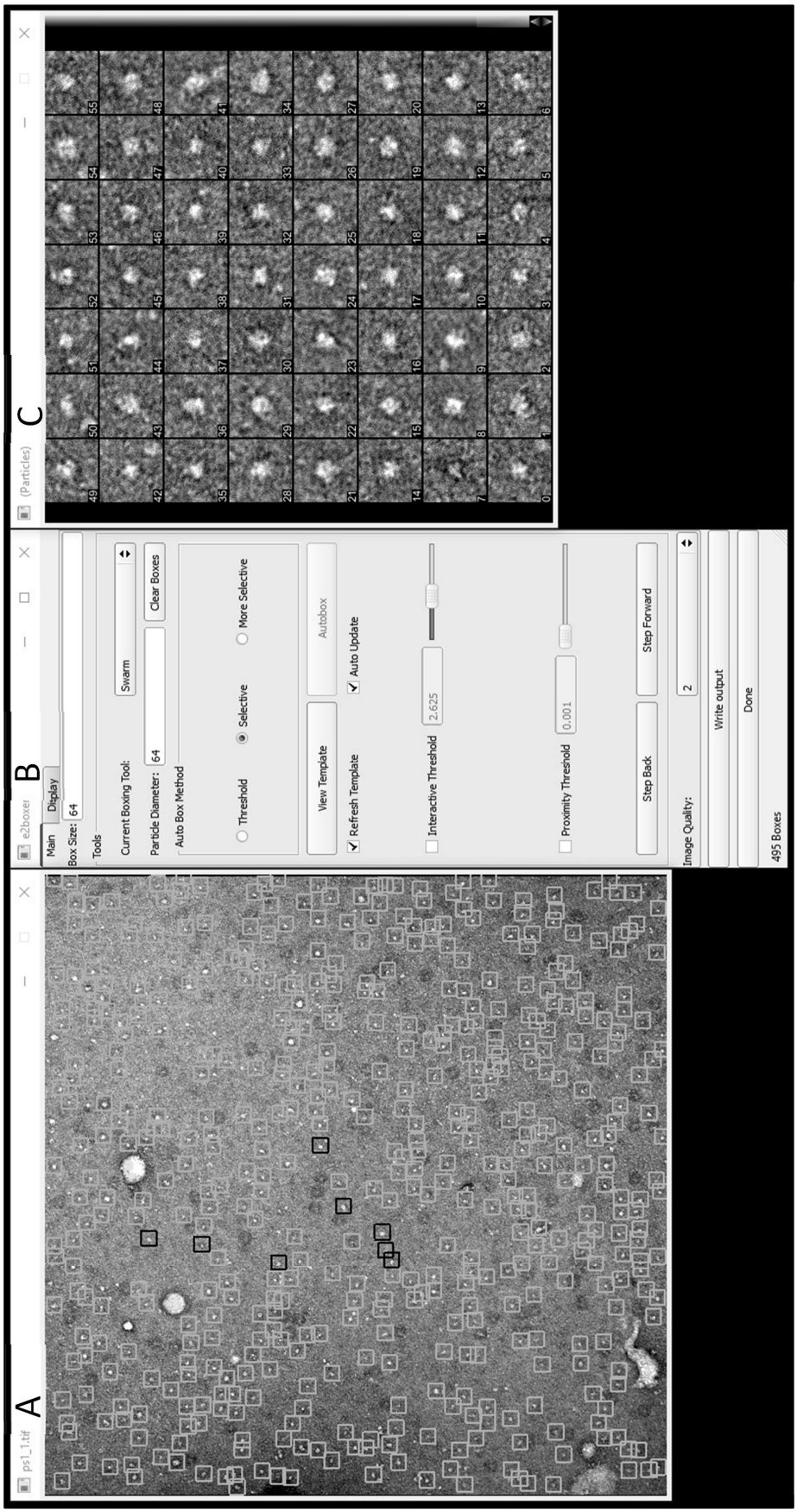


Figure 5.11 A screenshot for a typical single particle picking procedure by the 'boxer' module of EMAN2.

A) The original micrograph. B) EMAN2 'boxer' command window. C) The magnified boxed single particles.

From the particle boxing procedure, a data set of 13,341 individual protein complex images data set was built. Subsequently, the Imagic-5 software environment was used for image normalisation, band-pass filtering, reference-free alignment, and multi-variate statistical classification of the data set.

The boxed 13,341 single particle data were then subjected to reference-free alignment and multivariate statistical classification using Imagic-5 software. 2,024 particles were considered to be junk and removed, this leaving 11,317 particles being further processed. 100 characteristic views (class averages) were obtained after 4 rounds of iterative refinement, and reordered into oligomeric order in Figure 5.12. It can be clearly noticed that the single particles of Hik2 are present in various oligomeric states, from monomer to tetramer/double-trimer, and also some ambiguous ones that may represent broken Hik2 complexes.

When imaging at 80,000 × magnification, the side of each box, within which individual characteristic views were floated, represents 64 pixels that total 382 Å in length at the sampling frequency of 5.962 Å per pixel. Thus, a scale bar and protein size can be calculated and shown. The typical monomer, such as the one in position 10 in Figure 5.12, was measured to be ~ 6 nm in diameter, while the largest view, such as the one in position 66, being ~ 16 to 18 nm, both sizing within a reasonable range as discussed by (Erickson 2009).

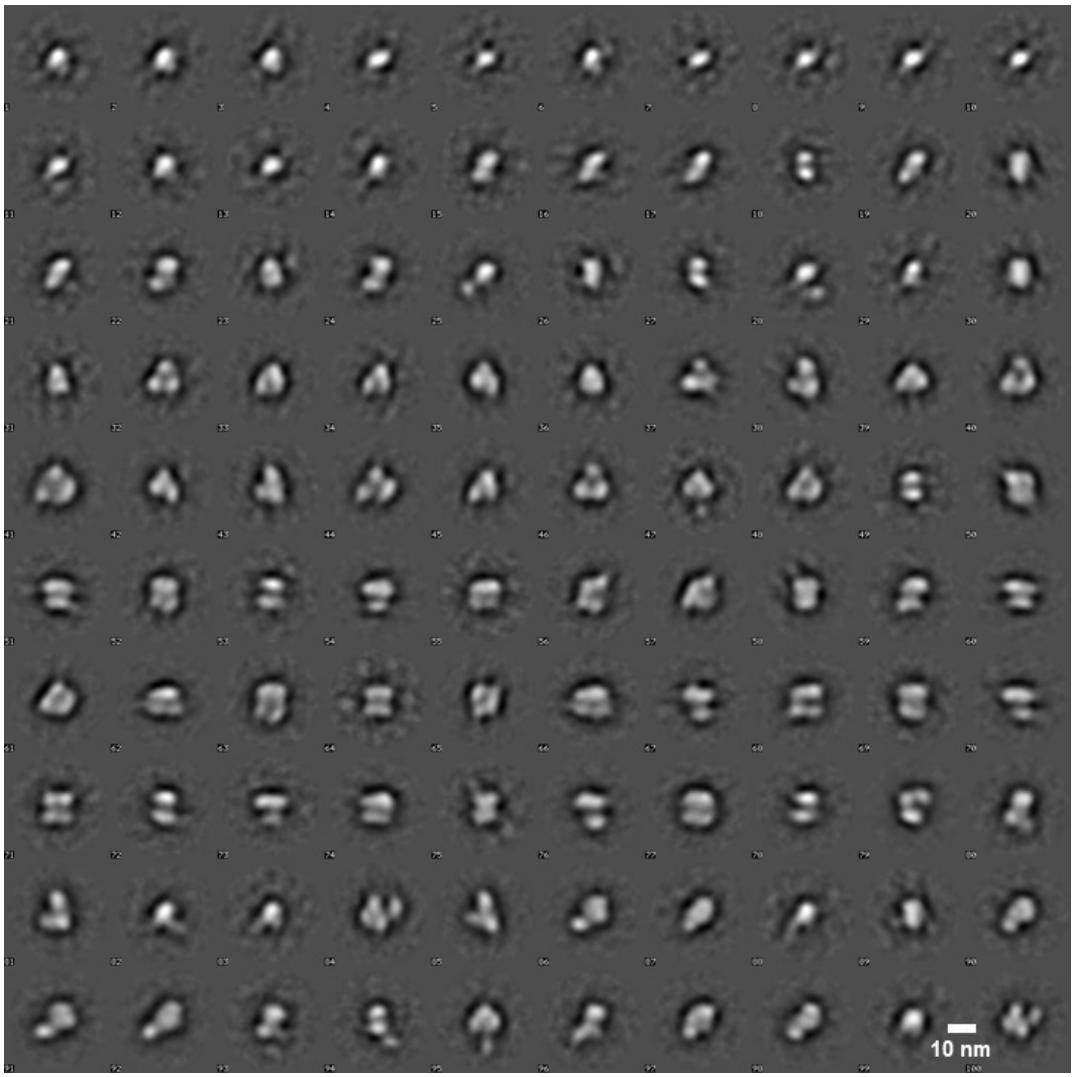
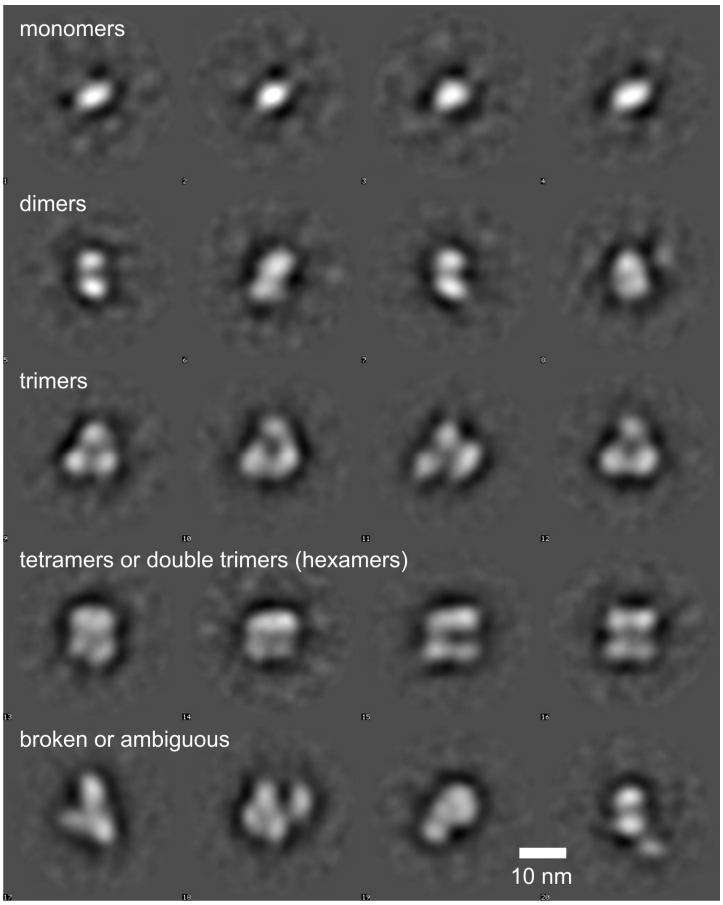


Figure 5.12 100 class averages of Hik2 single particles. These averages were averaged down from 11,317 Hik2 single particles after 4 rounds of iterative refinements. The scale bar represents 10 nm. Figure supplied by Dr. Nield.

4 characteristic class averages were selected subjectively from the 100 averages displayed in Figure 5.12, to represent the different oligomeric complex families: monomers, dimers, trimers, double dimers/hexamers, and broken or ambiguous density (Figure 5.13).

A



B

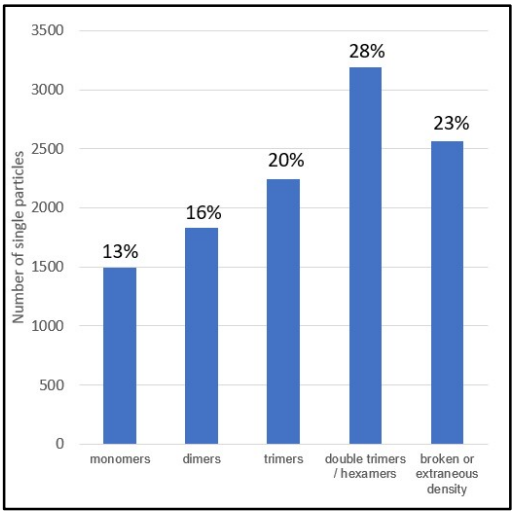


Figure 5.13 5 typical subpopulations selected from the 100 averages. A) 4 representative averages for each family attributed to oligomeric states of monomers, dimers, trimers, tetramers or potential double trimers in side elevation, and the remaining family of broken complexes, or those with extraneous density visible. Figure supplied by Dr. Nield. B) The histogram for each oligomeric state.

5.3 Discussion and future work

5.3.1 TEM-derived single particles are projections of a 3D object with different orientations

The TEM images are the projections of a 3D object (Figure 5.14), and the principle of single particle analysis is based on the 3D reprojections from 2D averages. Since the biological samples are highly radiation sensitive, low electron doses should be used when taking images, resulting in a high ratio of noise relative to the signal. Thus, a certain number of images need to be obtained and enough quantity of single particles should be picked (no less than 10,000 single particles for a homogeneous sample), for the alignment of several similar particles to each other and averaging them, to get a higher signal to noise ratio.

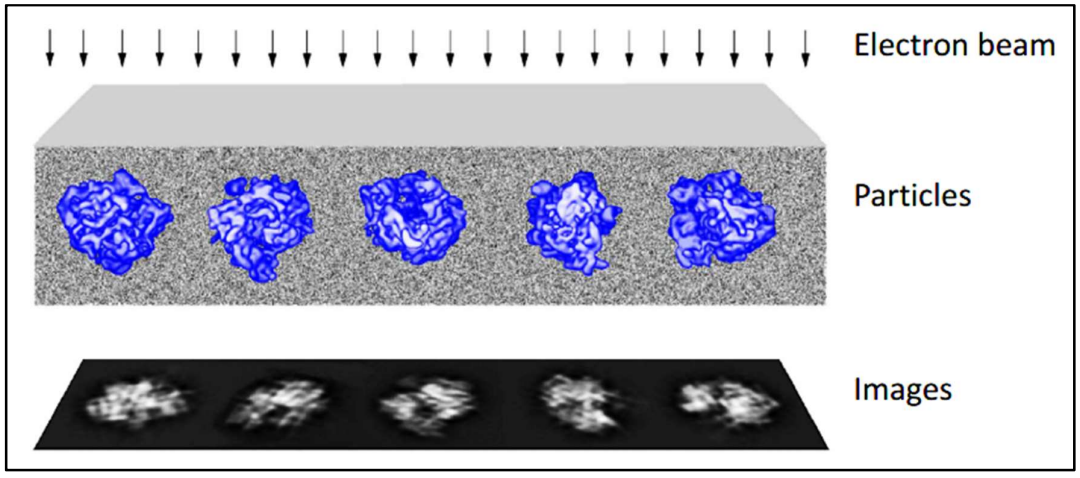


Figure 5.14 An example of 2D images derived from projections of a 3D shape in different orientations. Figure from Joachim Frank (http://biomachina.org/courses/structures/091.pdf).

When embedded on the grid surface, the particles are positioned with different, unknown orientations, moreover, the 2D averages do not have any depth, which makes

it difficult to tell the exact oligomeric states of Hik2 from Figure 5.13 A. For example, a dimer in lane 2 could be a tetramer in lane 4 that stands vertically to the grid, and there is the possibility that the trimers in lane 3 are actually a projection from another orientation of hexamers with a 3, 2-symmetry.

In order to get a 3D reconstruction of Hik2 at different oligomeric states, a much larger number of high quality micrographs need to be obtained, and single particles that might be in the same order oligomeric state should be classified separately, followed by the 3D reconstruction procedure for each possible class sum. Further biochemical experiments need to be done to determine if trimers actually exist, or if the trimers observed in Figure 5.13 A are the projections of hexamers; as well as investigating if the tetramers are dimers of dimers, or tetramers of four monomers. It will provide more information of the symmetry of the protein complexes to help with the 3D reconstruction.

5.3.2 Radiation damage and image drift

During imaging under TEM, the biological specimens, such as protein samples, are quite sensitive to bombardment by the incident electrons, and it is one of the limitations of acquiring high resolution micrographs. The electron-beam damage to the specimen are caused by the inelastic collisions, highly reactive ions and free radicals, which might disrupt chemical bonds, resulting in a loss of short-range order or radiolytic decomposition of the specimen, and initially losing the internal information of a particle (Egerton 2013). The pure black dots in the background of the image in Figure 5.8 B display the observable appearance of radiation damage.

The radiation damage in TEM cannot be eliminated, but various procedures may be applied to minimise it, such as using the low-dose technique to avoid pre-irradiation of the specimen region for later recording, and reducing the incident beam current density to some extent, as discussed in (Egerton 2013).

Another limitation of acquiring TEM micrographs with high resolution is the drift. When taking TEM images, the grid tends to move slightly under the electron beam, causing a drift, which will blur the image or introduce unwanted artefacts. The reasons causing drifts includes thermal expansion of sample, mechanical vibration, thermal expansion of microscope, insufficiently fixed sample, or inexperienced operation. Thus, the mechanical stability of the TEM and specimen, and the proficiency are important for the spatial resolution of the images.

Certainly, in this project, part of the TEM micrographs contain drift, as in Figure 5.12, where after a few rounds of averaging drift can be observed within the characteristic views, such as the ones in position 7, 8, 9 and 10, causing a loss of resolution, however, within the outcome of obtaining information in the 20 Å to 30 Å resolution range, this level of drift, while needed to be improved, is acceptable.

5.3.3 Oligomeric state affected by negative staining

It has been reported that the use of negative stain may affect the oligomeric state of proteins, either partially dissociate the oligomers, such as KLH (Keyhole limpet hemocyanin), from large homo-oligomeric proteins into decamers (De Carlo and Harris 2011) (Figure 5.15), or completely dissociate the protein complex, such as NtrC (De Carlo et al. 2006).

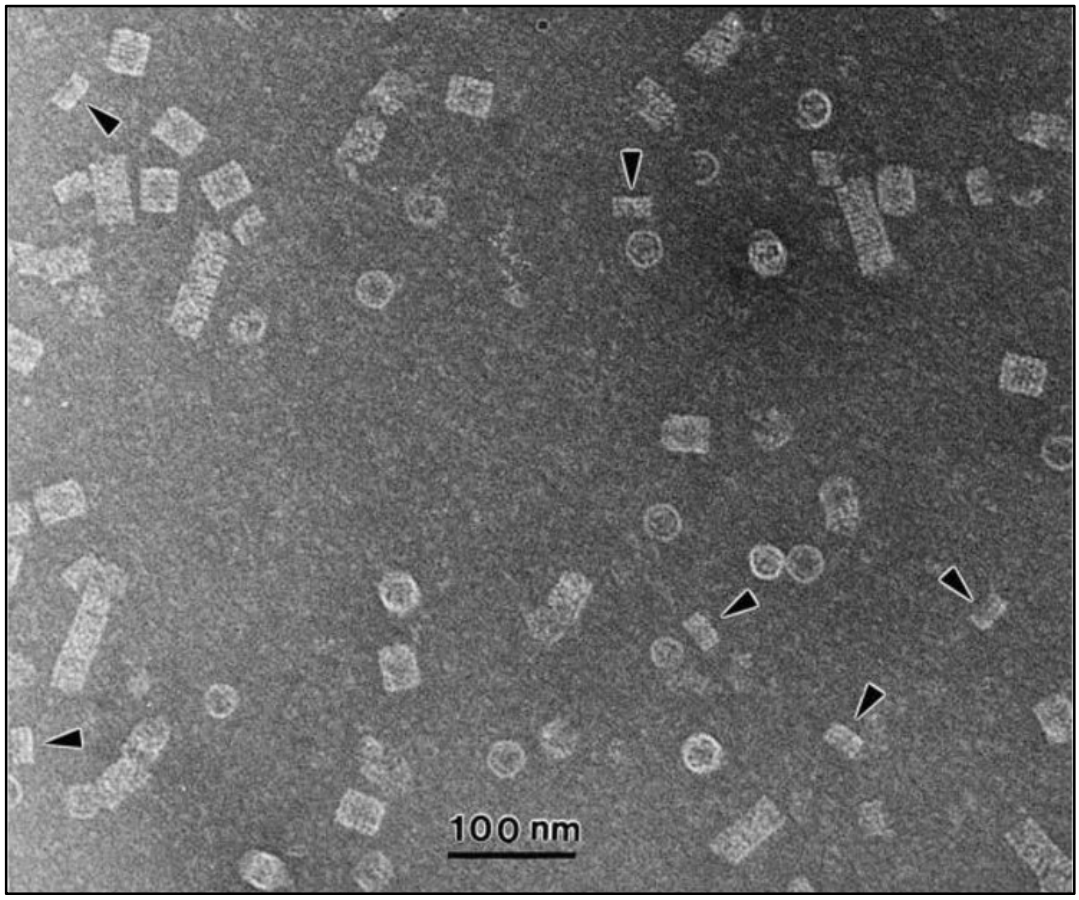


Figure 5.15 A TEM image of KLH sample prepared by negative staining. Increased exposure of the sample to the saturated ammonium molybdate staining solution (60 s) may induce dissociation of large homo-oligomeric proteins into decamers (indicated by arrowheads). Figure from (De Carlo and Harris 2011).

Low salt buffer was used in TEM sessions in this project to minimize the background noise caused by salt, and it is also the preferred condition for Hik2 activity. However, Hik2 protein sample was exposed to the concentrated PTA or UA when negatively stained, and it is possible the higher order oligomers observed do dissociate to some extent. Although it was also reported that UA might fix certain oligomeric protein structures in as few as 10 milliseconds (Zhao and Craig 2003), there might still be some effects due to the UA stain, or the interaction between the protein and the carbon-coated

grid, being the reason that the histogram in Figure 5.13 B is not consistent with our assumption for the oligomeric states of Hik2 under low salt conditions. One possible solution is to use the cryo-TEM technique which requires no prior staining procedures, and which may help to maintain the native oligomeric states of Hik2 complexes.

5.4 Conclusions

To date, there is no known structure for a dimeric full-length histidine kinase, either through lack of crystallisation or given that they are too large for NMR spectroscopy. Here, by combining TEM imaging technique with single particle computer analysis, a 3D map of a higher order oligomeric *Synechocystis* Hik2, and several 2D averages for its different oligomeric forms have been revealed to give us a first direct structural visualisation for the characterisation of Hik2. Problems overcome during TEM specimen preparation and the TEM imaging process were also discussed. Moreover, it may be anticipated that the further understanding of the protein oligomer formation will be obtained if, in the future, 3D map for each oligomeric form, can be calculated from TEM data.

Chapter 6 **Overall summary and future outlook**

As described in Chapter 1, section 1.2.1, there are Hik2 homologues present in all sequenced cyanobacterial species, and it is known that *hik2* mutants cannot survive (Paithoonrangsarid et al. 2004), thus Hik2 must play an important role in cyanobacteria. The specific homologue of Hik2 in plants identified as Chloroplast Sensor Kinase (CSK) has been discovered to function in connecting photosynthetic electron transport with chloroplast gene transcription (see Chapter 1 section 1.3.3). Both Hik2 and CSK have been determined to be histidine kinases. It is therefore important to characterise their structural information in order to understand how they carry out their functions in any signalling transduction pathway.

The oligomeric state for histidine kinase is essential for its autophosphorylation activity, since it has been suggested that most histidine kinases form dimers to catalyse the phosphorylation reaction on itself, or on a partner monomer via *cis*- or *trans*-autophosphorylation (Casino et al. 2014). However, there are also exceptions, for example, a blue light-activated histidine kinase EL346 in *Erythrobacter litoralis* HTCC2594 functions as a monomer. EL346 consists of a Light-Oxygen-Voltage (LOV) domain as its sensor domain, which binds its DHp-like domain and inhibits the activity by preventing the dimerisation in the dark. When placed in blue light, this LOV domain affects the interface of the DHp-like domain and CA domain, and initiates the phosphorylation activity (Rivera-Cancel et al. 2014). There is also a histidine kinase existing in tetrameric form, a membranous sensor kinase DcuS, which belongs to the dicarboxylate uptake sensor and regulator system in *E. coli*. In addition to its dimeric state, DcuS is also present in the tetrameric state in bacterial membranes and proteoliposomes, but it is not yet clear which form is functional (Scheu et al. 2010).

Moreover, as discussed in Chapter 3 section 3.3.5, the soluble histidine kinase ExsG in *Rhizobium* NT-26 exists as a hexamer with autophosphorylation activity.

Although it has been proved in this work and in the literature by various experiments, such as mass spectrometry, analytical ultracentrifugation, chemical cross-linking, fluorescence resonance energy transfer spectroscopy, either *in vitro* or *in vivo*, that histidine kinase can exist in different oligomeric forms, in addition it is the first-time direct structural visualisation of these oligomers has been made in this work (see Chapter 5, TEM studies). One 3D structure of a higher order oligomeric form of *Synechocystis* Hik2 was calculated (see Chapter 5, section 5.2.2), and several 2D averages for different oligomeric forms were revealed (see Chapter 5, section 5.2.3), displaying many varieties in Hik2 oligomeric state, such as the monomer, dimer, trimer (or a side view of hexamer), and the tetramer. However, the 3D model does not show an obvious elongated structure or a symmetry, possibly due to the lack of particle quantity or the low resolution; and we cannot clearly tell the oligomeric forms in its constituent 2D projections. More single particles in the overall dataset would improve the TEM work, even better a move into cryo-EM in order to avoid the effect of the negative staining. Attempts could be made to calculate 3D maps for each oligomeric state based on the dimension if over 10,000 single particles could be obtained for each.

In Chapter 3, it was observed from size exclusion chromatography, chemical cross-linking, and dynamic light scarring results that NaCl may dissociate the higher order oligomers of Hik2, possibly from hexamers into tetramers, and the DHp domain, which is usually responsible for the dimerisation and phosphor-accepting function, may be essential for the NaCl sensing; meanwhile Na^+ ion has been proved to inhibit the

autophosphorylation activity of Hik2 (Ibrahim et al. 2016). Thus, the autophosphorylation activity of Hik2 must strongly be related to the oligomeric state. It has been observed that the dissociation of higher order oligomers of ExsG is caused by increasing the concentration of ammonium acetate during native ion-mobility mass spectrometry experiments (Wojnowska et al. 2013). However, these authors proposed a model of the active 'open/relaxed' and inactive 'closed/compact' conformations of the hexameric ExsG, regulated by the phosphorylation state of its receiver domain, while it has been assumed in this thesis here, in Hik2, the DHp domain binds Na^+, resulting in an equilibrium shift of different oligomeric forms, and this also causes a conformational change of the tetrameric Hik2 between the active and inactive forms.

As demonstrated in Chapter 4, the sensor domain of Hik2 is quite insoluble, however, the full-length and the kinase domain of Hik2 were found to be soluble, which means it is likely that the sensor domain bends structurally towards the kinase domain so that the hydrophobic side of the surface can bury itself inside the protein. This would make the full-length protein relatively compact. However, the high non-ideality showed in the AUC results indicated the Hik2 oligomers to be somewhat elongated (see Chapter 3, section 3.2.5). In addition, the non-ideal behaviour may suggest oligomers of the same mass may not be in the same conformation, since they may have slightly different frictional force when sedimenting in aqueous solution due to their surface area differences.

Figure 6.1 proposes a model of the equilibrium between Hik2 oligomers in aqueous solution, and the Na^+ effect on the dissociation of the higher order oligomers. Under low salt conditions, Hik2 is autokinase active and in equilibrium between monomer, dimer,

tetramer, and hexamer forms. It is able to transfer phosphoryl groups to its cognate response regulators Rre1 and RppA (Ibrahim et al. 2016), in response to Cl^-, or redox signals, or other possible signals. In the presence of Na^+, the active tetramer tends towards another conformational state, which cannot bind and hydrolyse ATP, and this can be reversed when removing Na^+. In the meantime, the hexamers also have a tendency to dissociate to dimers, which would likely go on to form more tetramers. Additionally, it cannot yet be ruled out that the monomers may form inactive tetramers under high salt conditions.

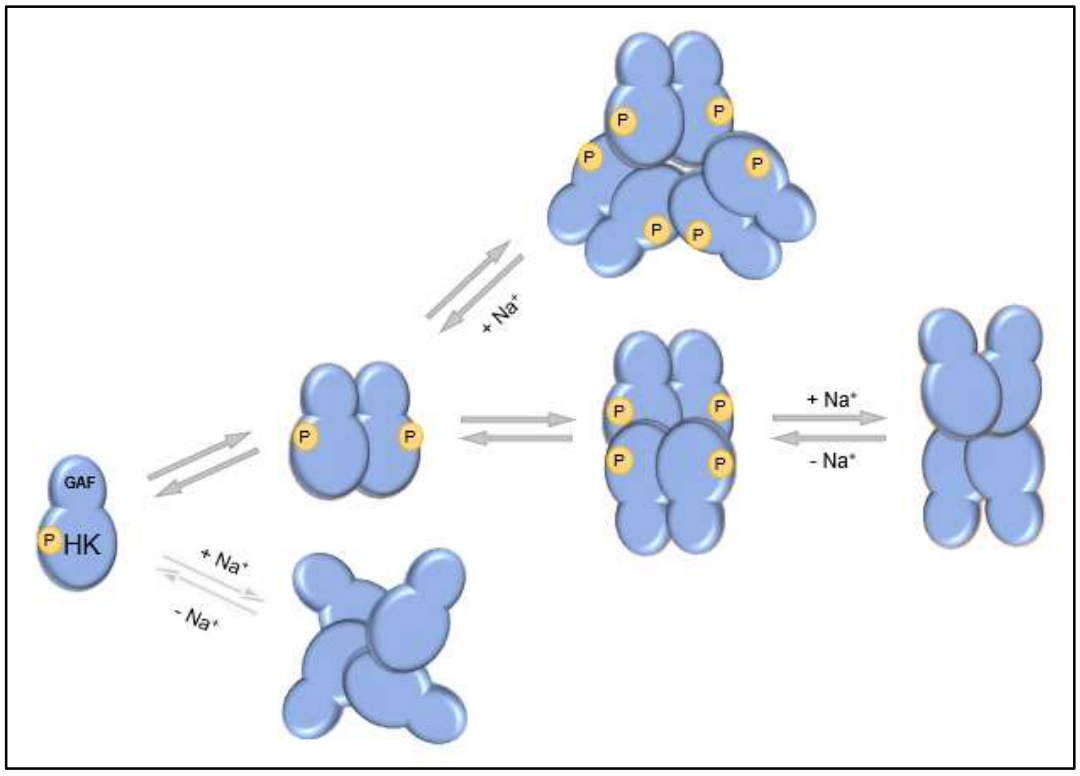


Figure 6.1 A proposed model for the equilibrium between Hik2 oligomers and the Na^+ effect on the dissociation. Under low salt conditions, Hik2 is autokinase active and in equilibrium between monomer, dimer, tetramer, and hexamer. In the presence of Na^+, the hexamers may dissociate to dimers, and the active tetramer form tends to become inactive in conformation, which can be reversed by removing Na^+. It is also possible that the monomers may form inactive tetramers directly under high salt conditions. GAF, the GAF domain; HK, histidine kinase domain; P, phosphoryl group.

In the future, more work should be done to study the dynamic equilibrium between different conformations, e.g. native ion-mobility mass spectrometry might be applied to probe different conformations by calculating the surface area of the protein complexes. Moreover, additional biochemical experiments such as combining phosphotransfer analysis with chemical cross-linking could be performed, in order to identify which oligomeric form(s) of Hik2 are able to transfer its phosphoryl group to its response regulators Rre1 and RppA. More effort might be put into crystallisation trials for the full-length or the sensor domain of Hik2 and CSK since the X-ray diffraction information would help to elucidate the signals sensing mechanism.

Histidine kinase is one component of the so-called two-component regulatory system, an important signal transduction pathway in bacteria, fungi and plants. This pathway allows those organisms to sense and respond to environmental changes, being absolutely essential for their adaptation and evolution. The structural characterisation of Hik2 in this thesis should prove crucial in aiding us to understand in greater detail the signal transduction mechanism of histidine kinases, providing information about the pathogenesis of relative diseases, potential drug targets, or for elucidating drug mechanisms of action to halt or terminate pathogenic bacteria (Bem et al. 2015), which in turn will relate to key protein kinases such as serine/threonine kinases and tyrosine kinases. For example, the acid-sensing histidine kinase ArsS regulates genes involved in acid resistance with its response regulator ArsR, this being a response to gastric acid stress in *Helicobacter pylori*, one of the risk factors causing gastric cancer. Studies on how ArsS weakens the response to gastric acid stress likely influences the progression of gastric carcinogenesis have been detailed (Krishna et al. 2016). Other recent studies also reviewed a variety of two-component system proteins, those involved in the

expression of virulence factors during infections from major pathogens of lung disease cystic fibrosis, and these proved that TCS inhibitors provided an option for treating bacterial infection (Velikova and Wellsa 2016). In addition, Hik2 and CSK have been suggested to play a key role in photosynthesis, one of the most important processes for all living organisms on Earth. Studies on these proteins will provide us with a deeper understanding of photosynthesis, leading to its optimisation, and thus eventually improve and optimise the crop agriculture and, tangentially, aid in the amelioration of the negative effects of global warming.

Appendix

Appendix 1

A

Protein standard	Kav	MW (Da)	logMW	Ve (ml)
Thyroglobulin	0.02903811	669000	5.82542612	7.95
Apoferritin	0.09437387	443000	5.64640373	9.03
Alcohol dehydrogenase	0.22444041	150000	5.17609126	11.18
Lysozyme	0.77253479	14300	4.15533604	20.24

Vo=7.47 ml

Vt=24.0 ml

Kav=(Ve-Vo)/(Vt-Vo)

B

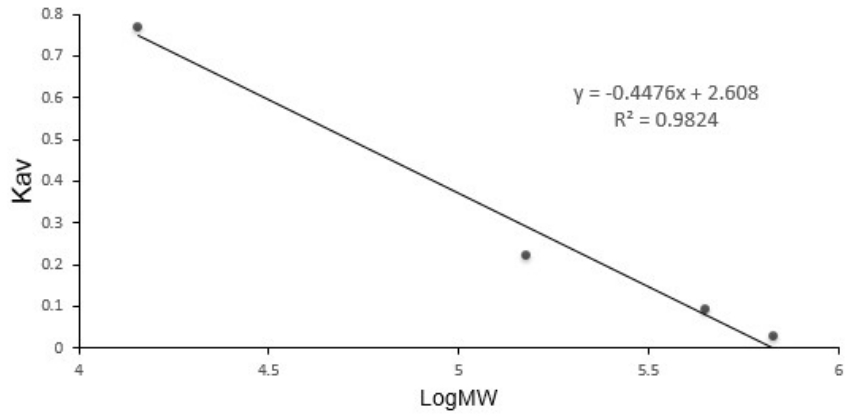


Figure S1 Calibration of Superdex 200 gel filtration column with low salt buffer. A) Data for protein standards used in calibration. B) The standard curve generated from the data in panel A. Low salt buffer contains 20 mM Tris-HCl, 10 mM NaCl (pH 7.5). Apparent molecular mass of Hik2 and the subdomain samples were calculated from the equation of the best-fit line.

Appendix 2

A

Protein standard	Kav	MW (Da)	LogMW	Ve (ml)
Thyroglobulin	0.05565638	669000	5.82542612	8.39
Apoferritin	0.14519056	443000	5.64640373	9.87
Alcohol dehydrogenase	0.28433152	150000	5.17609126	12.17
Lysozyme	0.69086509	14300	4.15533604	18.89

Vo = 7.47 ml

Vt = 24.0 ml

Kav = (Ve-Vo)/(Vt-Vo)

B

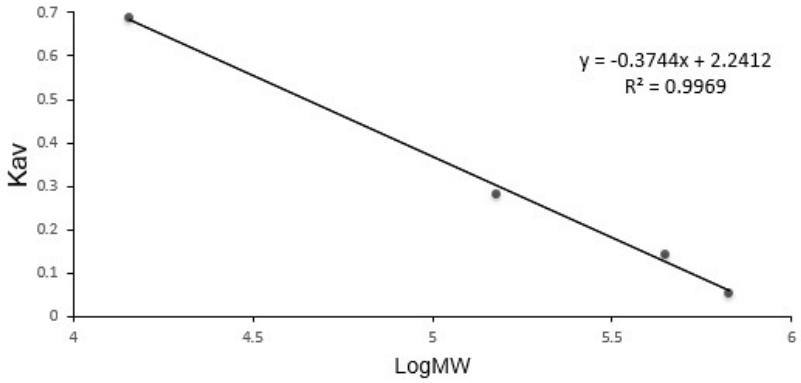


Figure S2 Calibration of Superdex 200 gel filtration column with high salt buffer. A) Data for protein standards used in calibration. B) The standard curve generated from the data in panel A. Low salt buffer contains 20 mM Tris-HCl, 500 mM NaCl (pH 7.5). Apparent molecular mass of Hik2 and the subdomain samples were calculated from the equation of the best-fit line.

Appendix 3

Figure S3 shows 600 characteristic views (class averages) of *Synechocystis* Hik2, obtained by relaxing the classification constraints, averaged down from 11,371 automatically picked single particles from 41 micrographs. The structural heterogeneity of Hik2 can be more fully observed from this figure.

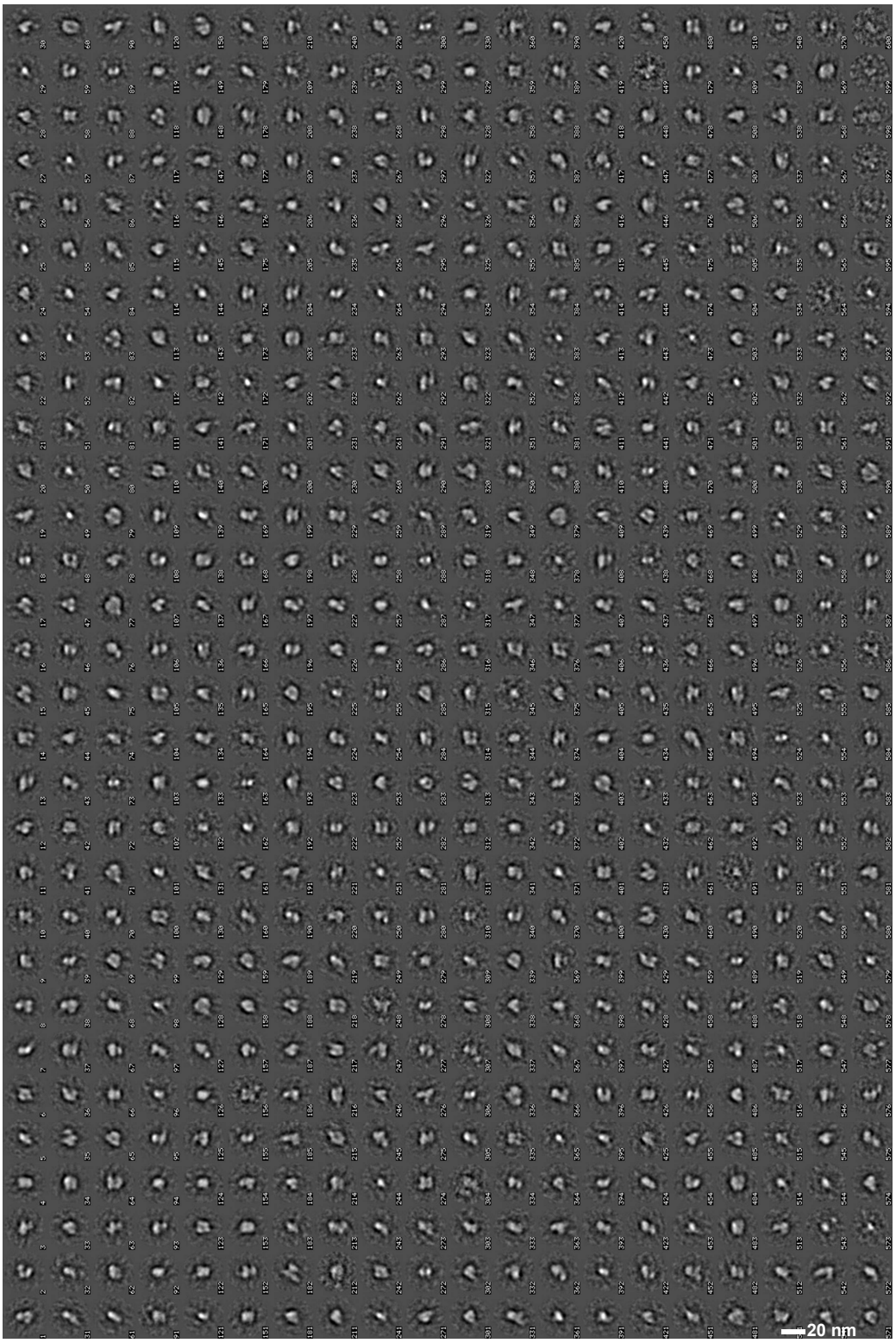


Figure S3 600 characteristic views (class averages) that averaged down from 11,371 single particles. Each boxed side of a single characteristic view represents 382 Å.

Appendix 4

Table S1 exhibits the single particle image processing statistics relating to the characteristic views (averages) presented in Figure 5.13 A, such locations cross-referenced with Figure 5.12 (100 averages). Each number below refers to the number of single particles present in each average. For example, the first monomer average in Figure 5.12 contains 107 particles.

Monomers	Dimers	Trimers	Double dimers / hexamers	Broken or extraneous density
107	82	106	86	289
133	111	152	164	94
129	138	137	83	112
124	109	197	95	107
95	88	102	78	98
127	128	213	182	90
85	129	111	47	122
90	96	164	68	141
121	125	117	202	121
99	117	119	181	91
87	102	168	92	128
106	94	102	169	127
81	117	143	124	123
107	85	165	161	129
	91	100	85	84
	117	57	47	74
	99	92	133	111
			67	76
			115	105
			115	140

			70	103
			114	99
			71	
			70	
			69	
			113	
			83	
			125	
			82	
			98	
14 averages	17 averages	17 averages	30 averages	22 averages
1491	1828	2245	3189	2564
11,317 classified particles 2,024 junks, removed 13,341 total				

Table S1 The single particle image processing statistics relating to the 100 averages presented in Figure 5.12.

References

Aiba, H., Nakasai, F., Mizushima, S. and Mizuno, T. (1989) Evidence for the physiological importance of the phosphotransfer between the two regulatory components, EnvZ and OmpR, in osmoregulation in Escherichia coli. *J Biol Chem* **264**(24): 14090-14094.

Albanesi, D., Martin, M., Trajtenberg, F., Mansilla, M.C., Haouz, A., Alzari, P.M., de Mendoza, D. and Buschiazzo, A. (2009) Structural plasticity and catalysis regulation of a thermosensor histidine kinase. *Proc Natl Acad Sci U S A* **106**(38): 16185-16190.

Allakhverdiev, S.I. and Murata, N. (2008) Salt stress inhibits photosystems II and I in cyanobacteria. *Photosynth Res* **98**(1-3): 529-539.

Allakhverdiev, S.I., Nishiyama, Y., Miyairi, S., Yamamoto, H., Inagaki, N., Kanesaki, Y. and Murata, N. (2002) Salt stress inhibits the repair of photodamaged photosystem II by suppressing the transcription and translation of psbA genes in synechocystis. *Plant Physiol* **130**(3): 1443-1453.

Allen, J.F. (1992) Protein phosphorylation in regulation of photosynthesis. *Biochim Biophys Acta* **1098**(3): 275-335.

Allen, J.F., Bennett, J., Steinback, K.E. and Arntzen, C.J. (1981) Chloroplast protein phosphorylation couples plastoquinone redox state to distribution of excitation energy between photosystems. *Nature* **291**(5810): 25-29.

Amos, L.A., Henderson, R. and Unwin, P.N. (1982) Three-dimensional structure determination by electron microscopy of two-dimensional crystals. *Prog Biophys Mol Biol* **39**(3): 183-231.

Anand, G.S., Goudreau, P.N. and Stock, A.M. (1998) Activation of methylesterase CheB: evidence of a dual role for the regulatory domain. *Biochemistry* **37**(40): 14038-14047.

Appleby, J.L., Parkinson, J.S. and Bourret, R.B. (1996) Signal transduction via the multi-step phosphorelay: not necessarily a road less traveled. *Cell* **86**(6): 845-848.

Aravind, L. and Ponting, C.P. (1997) The GAF domain: an evolutionary link between diverse phototransducing proteins. *Trends Biochem Sci* **22**(12): 458-459.

Aravind, L. and Ponting, C.P. (1999) The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins. *FEMS Microbiol Lett* **176**(1): 111-116.

Ashby, M.K. and Houmard, J. (2006) Cyanobacterial two-component proteins: structure, diversity, distribution, and evolution. *Microbiol Mol Biol Rev* **70**(2): 472-509.

Ashby, M.K. and Mullineaux, C.W. (1999) Cyanobacterial ycf27 gene products regulate energy transfer from phycobilisomes to photosystems I and II. *FEMS Microbiol Lett* **181**(2): 253-260.

Avila-Perez, M., Hellingwerf, K.J. and Kort, R. (2006) Blue light activates the sigmaB-dependent stress response of Bacillus subtilis via YtvA. *J Bacteriol* **188**(17): 6411-6414.

Bachhawat, P., Swapna, G.V., Montelione, G.T. and Stock, A.M. (2005) Mechanism of activation for transcription factor PhoB suggested by different modes of dimerization in the inactive and active states. *Structure* **13**(9): 1353-1363.

Baena-Gonzalez, E., Baginsky, S., Mulo, P., Summer, H., Aro, E.M. and Link, G. (2001) Chloroplast transcription at different light intensities. Glutathione-mediated phosphorylation of the major RNA polymerase involved in redox-regulated organellar gene expression. *Plant Physiol* **127**(3): 1044-1052.

Baginsky, S., Tiller, K., Pfannschmidt, T. and Link, G. (1999) PTK, the chloroplast RNA polymerase-associated protein kinase from mustard (Sinapis alba), mediates redox control of plastid in vitro transcription. *Plant Mol Biol* **39**(5): 1013-1023.

Bem, A.E., Velikova, N., Pellicer, M.T., Baarlen, P., Marina, A. and Wells, J.M. (2015) Bacterial histidine kinases as novel antibacterial drug targets. *ACS Chem Biol* **10**(1): 213-224.

Benesch, J.L. and Ruotolo, B.T. (2011) Mass spectrometry: come of age for structural and dynamical biology. *Curr Opin Struct Biol* **21**(5): 641-649.

Berman, E.S., Fortson, S.L., Checchi, K.D., Wu, L., Felton, J.S., Wu, K.J. and Kulp, K.S. (2008) Preparation of single cells for imaging/profiling mass spectrometry. *J Am Soc Mass Spectrom* **19**(8): 1230-1236.

Bhoo, S.H., Davis, S.J., Walker, J., Karniol, B. and Vierstra, R.D. (2001) Bacteriophytochromes are photochromic histidine kinases using a biliverdin chromophore. *Nature* **414**(6865): 776-779.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Gallo Cassarino, T., Bertoni, M., Bordoli, L. and Schwede, T. (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* **42**(Web Server issue): W252-258.

Bilwes, A.M., Alex, L.A., Crane, B.R. and Simon, M.I. (1999) Structure of CheA, a signal-transducing histidine kinase. *Cell* **96**(1): 131-141.

Bilwes, A.M., Quezada, C.M., Croal, L.R., Crane, B.R. and Simon, M.I. (2001) Nucleotide binding by the histidine kinase CheA. *Nat Struct Biol* **8**(4): 353-360.

Birck, C., Mourey, L., Gouet, P., Fabry, B., Schumacher, J., Rousseau, P., Kahn, D. and Samama, J.P. (1999) Conformational changes induced by phosphorylation of the FixJ receiver domain. *Structure* **7**(12): 1505-1515.

Blanco, A.G., Sola, M., Gomis-Ruth, F.X. and Coll, M. (2002) Tandem DNA recognition by PhoB, a two-component signal transduction transcriptional activator. *Structure* **10**(5): 701-713.

Blankenship, R.E. (2014) Molecular Mechanisms of Photosynthesis, *Wiley-Blackwell*.

Bogel, G., Schrempf, H. and Ortiz de Orue Lucana, D. (2009) The heme-binding protein HbpS regulates the activity of the Streptomyces reticuli iron-sensing histidine kinase SenS in a redox-dependent manner. *Amino Acids* **37**(4): 681-691.

Bourret, R.B. (2010) Receiver domain structure and function in response regulator proteins. *Curr Opin Microbiol* **13**(2): 142-149.

Brautigam, C.A., Padrick, S.B. and Schuck, P. (2013) Multi-signal sedimentation velocity analysis with mass conservation for determining the stoichiometry of protein complexes. *PLoS One* **8**(5): e62694.

Bruder, S., Linder, J.U., Martinez, S.E., Zheng, N., Beavo, J.A. and Schultz, J.E. (2005) The cyanobacterial tandem GAF domains from the cyaB2 adenylyl cyclase signal via both cAMP-binding sites. *Proc Natl Acad Sci U S A* **102**(8): 3088-3092.

Burgie, E.S., Walker, J.M., Phillips, G.N., Jr. and Vierstra, R.D. (2013) A photo-labile thioether linkage to phycoviolobilin provides the foundation for the blue/green photocycles in DXCF-cyanobacteriochromes. *Structure* **21**(1): 88-97.

Campbell, I. (2012) Analytical centrifugation. Biophysical Techniques, *Oxford University Press*.

Carazo, J.M., Sorzano, C.O., Oton, J., Marabini, R. and Vargas, J. (2015) Three-dimensional reconstruction methods in Single Particle Analysis from transmission electron microscopy data. *Arch Biochem Biophys* **581**: 39-48.

Casino, P., Miguel-Romero, L. and Marina, A. (2014) Visualizing autophosphorylation in histidine kinases. *Nat Commun* **5**: 3258.

Casino, P., Rubio, V. and Marina, A. (2009) Structural insight into partner specificity and phosphoryl transfer in two-component signal transduction. *Cell* **139**(2): 325-336.

Casino, P., Rubio, V. and Marina, A. (2010) The mechanism of signal transduction by two-component systems. *Curr Opin Struct Biol* **20**(6): 763-771.

Chang, C., Tesar, C., Gu, M., Babnigg, G., Joachimiak, A., Pokkuluri, P.R., Szurmant, H. and Schiffer, M. (2010) Extracytoplasmic PAS-like domains are common in signal transduction proteins. *J Bacteriol* **192**(4): 1156-1159.

Cheung, J., Bingman, C.A., Reyngold, M., Hendrickson, W.A. and Waldburger, C.D. (2008) Crystal structure of a functional dimer of the PhoQ sensor domain. *J Biol Chem* **283**(20): 13762-13770.

Cheung, J. and Hendrickson, W.A. (2008) Crystal structures of C4-dicarboxylate ligand complexes with sensor domains of histidine kinases DcuS and DctB. *J Biol Chem* **283**(44): 30256-30265.

Cheung, J. and Hendrickson, W.A. (2009) Structural analysis of ligand stimulation of the histidine kinase NarX. *Structure* **17**(2): 190-201.

Cho, U.S., Bader, M.W., Amaya, M.F., Daley, M.E., Klevit, R.E., Miller, S.I. and Xu, W. (2006) Metal bridges between the PhoQ sensor domain and the membrane regulate transmembrane signaling. *J Mol Biol* **356**(5): 1193-1206.

Chowdhury, S., Carter, J., Rollins, M.F., Golden, S.M., Jackson, R.N., Hoffmann, C., Nosaka, L., Bondy-Denomy, J., Maxwell, K.L., Davidson, A.R., Fischer, E.R., Lander, G.C. and Wiedenheft, B. (2017) Structure Reveals Mechanisms of Viral Suppressors that Intercept a CRISPR RNA-Guided Surveillance Complex. *Cell* **169**(1): 47-57 e11.

Cole, J.L., Lary, J.W., T, P.M. and Laue, T.M. (2008) Analytical ultracentrifugation: sedimentation velocity and sedimentation equilibrium. *Methods Cell Biol* **84**: 143-179.

Comeau, D.E., Ikenaka, K., Tsung, K.L. and Inouye, M. (1985) Primary characterization of the protein products of the Escherichia coli ompB locus: structure and regulation of synthesis of the OmpR and EnvZ proteins. *J Bacteriol* **164**(2): 578-584.

David, M., Daveran, M.L., Batut, J., Dedieu, A., Domergue, O., Ghai, J., Hertig, C., Boistard, P. and Kahn, D. (1988) Cascade regulation of nif gene expression in Rhizobium meliloti. *Cell* **54**(5): 671-683.

Davis, S.J., Vener, A.V. and Vierstra, R.D. (1999) Bacteriophytochromes: phytochrome-like photoreceptors from nonphotosynthetic eubacteria. *Science* **286**(5449): 2517-2520.

De Carlo, S., Chen, B., Hoover, T.R., Kondrashkina, E., Nogales, E. and Nixon, B.T. (2006) The structural basis for regulated assembly and function of the transcriptional activator NtrC. *Genes Dev* **20**(11): 1485-1495.

De Carlo, S. and Harris, J.R. (2011) Negative staining and cryo-negative staining of macromolecules and viruses for TEM. *Micron* **42**(2): 117-131.

Deusch, O., Landan, G., Roettger, M., Gruenheit, N., Kowallik, K.V., Allen, J.F., Martin, W. and Dagan, T. (2008) Genes of cyanobacterial origin in plant nuclear genomes point to a heterocyst-forming plastid ancestor. *Mol Biol Evol* **25**(4): 748-761.

Dubey, B.N., Lori, C., Ozaki, S., Fucile, G., Plaza-Menacho, I., Jenal, U. and Schirmer, T. (2016) Cyclic di-GMP mediates a histidine kinase/phosphatase switch by noncovalent domain cross-linking. *Sci Adv* **2**(9): e1600823.

Duysens, L.N.M. and Amesz, J. (1962) Function and identification of two photochemical systems in photosynthesis. *Biochimica et Biophysica Acta* **64**(2): 243-260.

Egerton, R.F. (2013) Control of radiation damage in the TEM. *Ultramicroscopy* **127**: 100-108.

Erickson, H.P. (2009) Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. *Biol Proced Online* **11**: 32-51.

Ferris, H.U., Coles, M., Lupas, A.N. and Hartmann, M.D. (2014) Crystallographic snapshot of the Escherichia coli EnvZ histidine kinase in an active conformation. *J Struct Biol* **186**(3): 376-379.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J. and Bateman, A. (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **44**(D1): D279-285.

Fischer, N., Neumann, P., Konevega, A.L., Bock, L.V., Ficner, R., Rodnina, M.V. and Stark, H. (2015) Structure of the E. coli ribosome-EF-Tu complex at <3 A resolution by Cs-corrected cryo-EM. *Nature* **520**(7548): 567-570.

Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M. and Leith, A. (1996) SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J Struct Biol* **116**(1): 190-199.

Galperin, M.Y. (2010) Diversity of structure and function of response regulator output domains. *Curr Opin Microbiol* **13**(2): 150-159.

Gantt, E. (1994) Supramolecular Membrane Organization. THE MOLECULAR BIOLOGY OF CYANOBACTERIA. D. A. Bryant**:** 119-138.

Gao, R. and Stock, A.M. (2009) Biological insights from structures of two-component proteins. *Annu Rev Microbiol* **63**: 133-154.

Garcia Vescovi, E., Soncini, F.C. and Groisman, E.A. (1996) Mg2+ as an extracellular signal: environmental regulation of Salmonella virulence. *Cell* **84**(1): 165-174.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D. and Bairoch, A. (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* **31**(13): 3784-3788.

Gavriilidou, A.F., Gulbakan, B. and Zenobi, R. (2015) Influence of Ammonium Acetate Concentration on Receptor-Ligand Binding Affinities Measured by Native Nano ESI-MS: A Systematic Study. *Anal Chem* **87**(20): 10378-10384.

Gest, H. (2002) History of the word photosynthesis and evolution of its definition. *Photosynth Res* **73**(1-3): 7-10.

Glazer, A.N. (1985) Light harvesting by phycobilisomes. *Annu Rev Biophys Biophys Chem* **14**: 47-77.

Golby, P., Davies, S., Kelly, D.J., Guest, J.R. and Andrews, S.C. (1999) Identification and characterization of a two-component sensor-kinase and response-regulator system (DcuS-DcuR) controlling gene expression in response to C4-dicarboxylates in Escherichia coli. *J Bacteriol* **181**(4): 1238-1248.

Gong, W., Hao, B., Mansy, S.S., Gonzalez, G., Gilles-Gonzalez, M.A. and Chan, M.K. (1998) Structure of a biological oxygen sensor: a new mechanism for heme-driven signal transduction. *Proc Natl Acad Sci U S A* **95**(26): 15177-15182.

Govindjee, R.G. (1975) Bioenergetics of photosynthesis, *Academic press*.

Green, M.R. and Sambrook, J. (2012) Isolation and Quantification of DNA. Molecular Cloning: A Laboratory Manual, Fourth Edition *Cold Spring Harbor Press*.

Greenfield, N.J. (2006) Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc* **1**(6): 2876-2890.

Hagemann, M. (2011) Molecular biology of cyanobacterial salt acclimation. *FEMS Microbiol Rev* **35**(1): 87-123.

Harauz, G., Borland, L., Bahr, G.F., Zeitler, E. and van Heel, M. (1987) Three-dimensional reconstruction of a human metaphase chromosome from electron micrographs. *Chromosoma* **95**(5): 366-374.

Hefti, M.H., Francoijs, K.J., de Vries, S.C., Dixon, R. and Vervoort, J. (2004) The PAS fold. A redefinition of the PAS domain based upon structural prediction. *Eur J Biochem* **271**(6): 1198-1208.

Heintzen, C., Loros, J.J. and Dunlap, J.C. (2001) The PAS protein VIVID defines a clock-associated feedback loop that represses light input, modulates gating, and regulates clock resetting. *Cell* **104**(3): 453-464.

Heitmann, P. (1968) A model for sulfhydryl groups in proteins. Hydrophobic interactions of the cystein side chain in micelles. *Eur J Biochem* **3**(3): 346-350.

Huala, E., Oeller, P.W., Liscum, E., Han, I.S., Larsen, E. and Briggs, W.R. (1997) Arabidopsis NPH1: a protein kinase with a putative redox-sensing domain. *Science* **278**(5346): 2120-2123.

Hwang, I., Chen, H.C. and Sheen, J. (2002) Two-component signal transduction pathways in Arabidopsis. *Plant Physiol* **129**(2): 500-515.

Ibrahim, I.M., Puthiyaveetil, S. and Allen, J.F. (2016) A Two-Component Regulatory System in Transcriptional Control of Photosystem Stoichiometry: Redox-Dependent and Sodium Ion-Dependent Phosphoryl Transfer from Cyanobacterial Histidine Kinase Hik2 to Response Regulators Rre1 and RppA. *Front Plant Sci* **7**: 137.

Ibrahim, I.M., Puthiyaveetil, S., Khan, C. and Allen, J.F. (2016) Probing the nucleotide-binding activity of a redox sensor: two-component regulatory control in chloroplasts. *Photosynth Res* **130**(1-3): 93-101.

Ibrahim, I.M., Wang, L., Puthiyaveetil, S., Krauss, N., Nield, J. and Allen, J.F. (2017) Oligomeric states in sodium ion-dependent regulation of cyanobacterial histidine kinase-2. *Protoplasma*. https://doi.org/10.1007/s00709-017-1196-7.

Jenal, U. and Galperin, M.Y. (2009) Single domain response regulators: molecular switches with emerging roles in cell organization and dynamics. *Curr Opin Microbiol* **12**(2): 152-160.

Jing, X., Jaw, J., Robinson, H.H. and Schubot, F.D. (2010) Crystal structure and oligomeric state of the RetS signaling kinase sensory domain. *Proteins* **78**(7): 1631-1640.

Kanamaru, K., Fujiwara, M., Seki, M., Katagiri, T., Nakamura, M., Mochizuki, N., Nagatani, A., Shinozaki, K., Tanaka, K. and Takahashi, H. (1999) Plastidic RNA polymerase sigma factors in Arabidopsis. *Plant Cell Physiol* **40**(8): 832-842.

Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M. and Tabata, S. (1996) Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res* **3**(3): 109-136.

Kanesaki, Y., Shiwa, Y., Tajima, N., Suzuki, M., Watanabe, S., Sato, N., Ikeuchi, M. and Yoshikawa, H. (2012) Identification of substrain-specific mutations by

massively parallel whole-genome resequencing of Synechocystis sp. PCC 6803. *DNA Res* **19**(1): 67-79.

Kaspar, S., Perozzo, R., Reinelt, S., Meyer, M., Pfister, K., Scapozza, L. and Bott, M. (1999) The periplasmic domain of the histidine autokinase CitA functions as a highly specific citrate receptor. *Mol Microbiol* **33**(4): 858-872.

Kawakami, K., Umena, Y., Kamiya, N. and Shen, J.R. (2009) Location of chloride and its possible functions in oxygen-evolving photosystem II revealed by X-ray crystallography. *Proc Natl Acad Sci U S A* **106**(21): 8567-8572.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. and Sternberg, M.J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**(6): 845-858.

Kirkwood, K.J., Ahmad, Y., Larance, M. and Lamond, A.I. (2013) Characterization of native protein complexes and protein isoform variation using size-fractionation-based quantitative proteomics. *Mol Cell Proteomics* **12**(12): 3851-3873.

Kiselev, N.A., Sherman, M.B. and Tsuprun, V.L. (1990) Negative staining of proteins. *Electron Microsc Rev* **3**(1): 43-72.

Kisielowski, C., Freitag, B., Bischoff, M., van Lin, H., Lazar, S., Knippels, G., Tiemeijer, P., van der Stam, M., von Harrach, S., Stekelenburg, M., Haider, M., Uhlemann, S., Muller, H., Hartel, P., Kabius, B., Miller, D., Petrov, I., Olson, E.A., Donchev, T., Kenik, E.A., Lupini, A.R., Bentley, J., Pennycook, S.J., Anderson, I.M., Minor, A.M., Schmid, A.K., Duden, T., Radmilovic, V., Ramasse, Q.M., Watanabe, M., Erni, R., Stach, E.A., Denes, P. and Dahmen, U. (2008) Detection of single atoms and buried defects in three dimensions by aberration-corrected electron microscope with 0.5-A information limit. *Microsc Microanal* **14**(5): 469-477.

Kotajima, T., Shiraiwa, Y. and Suzuki, I. (2014) Functional analysis of the N-terminal region of an essential histidine kinase, Hik2, in the cyanobacterium Synechocystis sp. PCC 6803. *FEMS Microbiol Lett* **351**(1): 88-94.

Krishna, U., Romero-Gallo, J., Suarez, G., Azah, A., Krezel, A.M., Varga, M.G., Forsyth, M.H. and Peek, R.M., Jr. (2016) Genetic Evolution of a Helicobacter pylori Acid-Sensing Histidine Kinase and Gastric Disease. *J Infect Dis* **214**(4): 644-648.

Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**(3): 567-580.

Kumar, A., Toledo, J.C., Patel, R.P., Lancaster, J.R., Jr. and Steyn, A.J. (2007) Mycobacterium tuberculosis DosS is a redox sensor and DosT is a hypoxia sensor. *Proc Natl Acad Sci U S A* **104**(28): 11568-11573.

Kumar, S., Kateriya, S., Singh, V.S., Tanwar, M., Agarwal, S., Singh, H., Khurana, J.P., Amla, D.V. and Tripathi, A.K. (2012) Bacteriophytochrome controls carotenoid-independent response to photodynamic stress in a non-photosynthetic rhizobacterium, Azospirillum brasilense Sp7. *Sci Rep* **2**: 872.

Lambert, C., Leonard, N., De Bolle, X. and Depiereux, E. (2002) ESyPred3D: Prediction of proteins 3D structures. *Bioinformatics* **18**(9): 1250-1256.

Letunic, I., Doerks, T. and Bork, P. (2015) SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res* **43**(Database issue): D257-260.

Li, H. and Sherman, L.A. (2000) A redox-responsive regulator of photosynthesis gene expression in the cyanobacterium Synechocystis sp. Strain PCC 6803. *J Bacteriol* **182**(15): 4268-4277.

Li, H., Zhang, J., Vierstra, R.D. and Li, H. (2010) Quaternary organization of a phytochrome dimer as revealed by cryoelectron microscopy. *Proc Natl Acad Sci U S A* **107**(24): 10872-10877.

Liao, H.Y. and Frank, J. (2010) Definition and estimation of resolution in single-particle reconstructions. *Structure* **18**(7): 768-775.

Liao, Y. (2007) Practical Electron Microscopy and Database. *http://www.globalsino.com/EM/*.

Link, G. (2003) Redox regulation of chloroplast transcription. *Antioxid Redox Signal* **5**(1): 79-87.

Liu, L.N. (2016) Distribution and dynamics of electron transport complexes in cyanobacterial thylakoid membranes. *Biochim Biophys Acta* **1857**(3): 256-265.

Lohrmann, J. and Harter, K. (2002) Plant two-component signaling systems and the role of response regulators. *Plant Physiol* **128**(2): 363-369.

Los, D.A., Zorina, A., Sinetova, M., Kryazhov, S., Mironov, K. and Zinchenko, V.V. (2010) Stress sensors and signal transducers in cyanobacteria. *Sensors (Basel)* **10**(3): 2386-2415.

Ludtke, S.J. (2010) 3-D structures of macromolecules using single-particle analysis in EMAN. *Methods Mol Biol* **673**: 157-173.

Ludtke, S.J., Baldwin, P.R. and Chiu, W. (1999) EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* **128**(1): 82-97.

Maiti, R., Van Domselaar, G.H., Zhang, H. and Wishart, D.S. (2004) SuperPose: a simple server for sophisticated structural superposition. *Nucleic Acids Res* **32**(Web Server issue): W590-594.

Marina, A., Mott, C., Auyzenberg, A., Hendrickson, W.A. and Waldburger, C.D. (2001) Structural and mutational analysis of the PhoQ histidine kinase catalytic domain. Insight into the reaction mechanism. *J Biol Chem* **276**(44): 41182-41190.

Marina, A., Waldburger, C.D. and Hendrickson, W.A. (2005) Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein. *EMBO J* **24**(24): 4247-4259.

Maris, A.E., Sawaya, M.R., Kaczor-Grzeskowiak, M., Jarvis, M.R., Bearson, S.M., Kopka, M.L., Schroder, I., Gunsalus, R.P. and Dickerson, R.E. (2002) Dimerization allows DNA target site recognition by the NarL response regulator. *Nat Struct Biol* **9**(10): 771-778.

Markson, J.S., Piechura, J.R., Puszynska, A.M. and O'Shea, E.K. (2013) Circadian control of global gene expression by the cyanobacterial master regulator RpaA. *Cell* **155**(6): 1396-1408.

Martin, M., Albanesi, D., Alzari, P.M. and de Mendoza, D. (2009) Functional in vitro assembly of the integral membrane bacterial thermosensor DesK. *Protein Expr Purif* **66**(1): 39-45.

Martin, W. and Kowallik, K.V. (1999) Annotated English translation of Mereschkowsky's 1905 paper 'Über Natur und Ursprung der Chromatophoren im Pflanzenreiche'. *European Journal of Phycology* **34**(3): 287-295.

Martinez, S.E., Beavo, J.A. and Hol, W.G. (2002) GAF domains: two-billion-year-old molecular switches that bind cyclic nucleotides. *Mol Interv* **2**(5): 317-323.

Martinez, S.E., Bruder, S., Schultz, A., Zheng, N., Schultz, J.E., Beavo, J.A. and Linder, J.U. (2005) Crystal structure of the tandem GAF domains from a cyanobacterial adenylyl cyclase: modes of ligand binding and dimerization. *Proc Natl Acad Sci U S A* **102**(8): 3082-3087.

Mattison, K. and Kenney, L.J. (2002) Phosphorylation alters the interaction of the response regulator OmpR with its sensor kinase EnvZ. *J Biol Chem* **277**(13): 11143-11148.

Mielańczyk, L., Matysiak, N., Klymenko, O. and Wojnicz, R. (2015) Transmission Electron Microscopy of Biological Samples. The Transmission Electron Microscope - Theory and Applications. K. Maaz, *InTech*.

Minoda, A. and Tanaka, K. (2005) Roles of the transcription factors encoded in the plastid genome of Cyanidioschyzon merolae. *Photosynthesis: fundamental aspects to global perspectives (eds A. vd Est & D. Bruce)*: 728-729.

Mitchell, P. (1961) Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature* **191**: 144-148.

Moglich, A., Ayers, R.A. and Moffat, K. (2009) Structure and signaling mechanism of Per-ARNT-Sim domains. *Structure* **17**(10): 1282-1294.

Moglich, A. and Moffat, K. (2007) Structural basis for light-dependent signaling in the dimeric LOV domain of the photosensor YtvA. *J Mol Biol* **373**(1): 112-126.

Moore, J.O. and Hendrickson, W.A. (2009) Structural analysis of sensor domains from the TMAO-responsive histidine kinase receptor TorS. *Structure* **17**(9): 1195-1204.

Moore, J.O. and Hendrickson, W.A. (2012) An asymmetry-to-symmetry switch in signal transmission by the histidine kinase receptor for TMAO. *Structure* **20**(4): 729-741.

Mulekar, J.J. and Huq, E. (2014) Expanding roles of protein kinase CK2 in regulating plant growth and development. *J Exp Bot* **65**(11): 2883-2893.

Myasnikov, A.G., Kundhavai Natchiar, S., Nebout, M., Hazemann, I., Imbert, V., Khatter, H., Peyron, J.F. and Klaholz, B.P. (2016) Structure-function insights reveal the human ribosome as a cancer target for antibiotics. *Nat Commun* **7**: 12856.

Myers, J. (1971) Enhancement Studies in Photosynthesis. *Annual Review of Plant Physiology* **22**(1): 289-312.

Neiditch, M.B., Federle, M.J., Pompeani, A.J., Kelly, R.C., Swem, D.L., Jeffrey, P.D., Bassler, B.L. and Hughson, F.M. (2006) Ligand-induced asymmetry in histidine sensor kinase complex regulates quorum sensing. *Cell* **126**(6): 1095-1108.

Nikaido, H. and Vaara, M. (1985) Molecular basis of bacterial outer membrane permeability. *Microbiol Rev* **49**(1): 1-32.

Ninfa, A.J., Ballou, D.P. and Benore, M. (2009) Gel Electrophoresis of Proteins. Fundamental Laboratory Approaches for Biochemistry and Biotechnology, Second Edition, *John Wiley & Sons*.

Nixon, B.T., Ronson, C.W. and Ausubel, F.M. (1986) Two-component regulatory systems responsive to environmental stimuli share strongly conserved domains with the nitrogen assimilation regulatory genes ntrB and ntrC. *Proc Natl Acad Sci U S A* **83**(20): 7850-7854.

Ogrzewalla, K., Piotrowski, M., Reinbothe, S. and Link, G. (2002) The plastid transcription kinase from mustard (Sinapis alba L.). A nuclear-encoded CK2-type

chloroplast enzyme with redox-sensitive function. *Eur J Biochem* **269**(13): 3329-3337.

Ohnishi, N. and Murata, N. (2006) Glycinebetaine counteracts the inhibitory effects of salt stress on the degradation and synthesis of D1 protein during photoinhibition in Synechococcus sp. PCC 7942. *Plant Physiol* **141**(2): 758-765.

Oka, A., Sakai, H. and Iwakoshi, S. (2002) His-Asp phosphorelay signal transduction in higher plants: receptors and response regulators for cytokinin signaling in Arabidopsis thaliana. *Genes Genet Syst* **77**(6): 383-391.

Paithoonrangsarid, K., Shoumskaya, M.A., Kanesaki, Y., Satoh, S., Tabata, S., Los, D.A., Zinchenko, V.V., Hayashi, H., Tanticharoen, M., Suzuki, I. and Murata, N. (2004) Five histidine kinases perceive osmotic stress and regulate distinct sets of genes in Synechocystis. *J Biol Chem* **279**(51): 53078-53086.

Peter, A.P., Lakshmanan, K., Mohandass, S., Varadharaj, S., Thilagar, S., Abdul Kareem, K.A., Dharmar, P., Gopalakrishnan, S. and Lakshmanan, U. (2015) Cyanobacterial KnowledgeBase (CKB), a Compendium of Cyanobacterial Genomes and Proteomes. *PLoS One* **10**(8): e0136262.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**(13): 1605-1612.

Pfannschmidt, T., Nilsson, A., Tullberg, A., Link, G. and Allen, J.F. (1999) Direct transcriptional control of the chloroplast genes psbA and psaAB adjusts photosynthesis to light energy distribution in plants. *IUBMB Life* **48**(3): 271-276.

Podust, L.M., Ioanoviciu, A. and Ortiz de Montellano, P.R. (2008) 2.3 A X-ray structure of the heme-bound GAF domain of sensory histidine kinase DosT of Mycobacterium tuberculosis. *Biochemistry* **47**(47): 12523-12531.

Prilusky, J., Felder, C.E., Zeev-Ben-Mordehai, T., Rydberg, E.H., Man, O., Beckmann, J.S., Silman, I. and Sussman, J.L. (2005) FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* **21**(16): 3435-3438.

Prost, L.R. and Miller, S.I. (2008) The Salmonellae PhoQ sensor: mechanisms of detection of phagosome signals. *Cell Microbiol* **10**(3): 576-582.

Puigbo, P., Guzman, E., Romeu, A. and Garcia-Vallve, S. (2007) OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. *Nucleic Acids Res* **35**(Web Server issue): W126-131.

Puthiyaveetil, S. and Allen, J.F. (2008) Transients in chloroplast gene transcription. *Biochem Biophys Res Commun* **368**(4): 871-874.

Puthiyaveetil, S., Ibrahim, I.M. and Allen, J.F. (2013) Evolutionary rewiring: a modified prokaryotic gene-regulatory pathway in chloroplasts. *Philos Trans R Soc Lond B Biol Sci* **368**(1622): 20120260.

Puthiyaveetil, S., Ibrahim, I.M., Jelicic, B., Tomasic, A., Fulgosi, H. and Allen, J.F. (2010) Transcriptional control of photosynthesis genes: the evolutionarily conserved regulatory mechanism in plastid genome function. *Genome Biol Evol* **2**: 888-896.

Puthiyaveetil, S., Kavanagh, T.A., Cain, P., Sullivan, J.A., Newell, C.A., Gray, J.C., Robinson, C., van der Giezen, M., Rogers, M.B. and Allen, J.F. (2008) The ancestral symbiont sensor kinase CSK links photosynthesis with gene expression in chloroplasts. *Proc Natl Acad Sci U S A* **105**(29): 10061-10066.

Quinn, J., Malakasi, P., Smith, D.A., Cheetham, J., Buck, V., Millar, J.B. and Morgan, B.A. (2011) Two-component mediated peroxide sensing and signal transduction in fission yeast. *Antioxid Redox Signal* **15**(1): 153-165.

Rai, N., Nollmann, M., Spotorno, B., Tassara, G., Byron, O. and Rocco, M. (2005) SOMO (SOlution MOdeler) differences between X-Ray- and NMR-derived bead models suggest a role for side chain flexibility in protein hydrodynamics. *Structure* **13**(5): 723-734.

Reimer, L., Kohl, H. (2008) Transmission Electron Microscopy, *Springer-Verlag New York*.

Rivalta, I., Amin, M., Luber, S., Vassiliev, S., Pokhrel, R., Umena, Y., Kawakami, K., Shen, J.R., Kamiya, N., Bruce, D., Brudvig, G.W., Gunner, M.R. and Batista,

V.S. (2011) Structural-functional role of chloride in photosystem II. *Biochemistry* **50**(29): 6312-6315.

Rivera-Cancel, G., Ko, W.H., Tomchick, D.R., Correa, F. and Gardner, K.H. (2014) Full-length structure of a monomeric histidine kinase reveals basis for sensory regulation. *Proc Natl Acad Sci U S A* **111**(50): 17839-17844.

Roberts, D.M., Liao, R.P., Wisedchaisri, G., Hol, W.G. and Sherman, D.R. (2004) Two sensor kinases contribute to the hypoxic response of Mycobacterium tuberculosis. *J Biol Chem* **279**(22): 23082-23087.

Rohou, A. and Grigorieff, N. (2015) CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**(2): 216-221.

Ruprecht, J. and Nield, J. (2001) Determining the structure of biological macromolecules by transmission electron microscopy, single particle analysis and 3D reconstruction. *Prog Biophys Mol Biol* **75**(3): 121-164.

Sato, S., Shimoda, Y., Muraki, A., Kohara, M., Nakamura, Y. and Tabata, S. (2007) A large-scale protein protein interaction analysis in Synechocystis sp. PCC6803. *DNA Res* **14**(5): 207-216.

Saxton, W.O. and Baumeister, W. (1982) The correlation averaging of a regularly arranged bacterial cell envelope protein. *J Microsc* **127**(Pt 2): 127-138.

Scheres, S.H. (2012) RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180**(3): 519-530.

Scheu, P.D., Kim, O.B., Griesinger, C. and Unden, G. (2010) Sensing by the membrane-bound sensor kinase DcuS: exogenous versus endogenous sensing of C(4)-dicarboxylates in bacteria. *Future Microbiol* **5**(9): 1383-1402.

Scheu, P.D., Liao, Y.F., Bauer, J., Kneuper, H., Basche, T., Unden, G. and Erker, W. (2010) Oligomeric sensor kinase DcuS in the membrane of Escherichia coli and in proteoliposomes: chemical cross-linking and FRET spectroscopy. *J Bacteriol* **192**(13): 3474-3483.

Schmetterer, G., Alge, D. and Gregor, W. (1994) Deletion of cytochrome c oxidase genes from the cyanobacterium Synechocystis sp. PCC6803: Evidence for alternative respiratory pathways. *Photosynth Res* **42**(1): 43-50.

Schrader, J., Henneberg, F., Mata, R.A., Tittmann, K., Schneider, T.R., Stark, H., Bourenkov, G. and Chari, A. (2016) The inhibition mechanism of human 20S proteasomes enables next-generation inhibitor design. *Science* **353**(6299): 594-598.

Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J* **78**(3): 1606-1619.

Schultz, J., Milpetz, F., Bork, P. and Ponting, C.P. (1998) SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A* **95**(11): 5857-5864.

Sevvana, M., Vijayan, V., Zweckstetter, M., Reinelt, S., Madden, D.R., Herbst-Irmer, R., Sheldrick, G.M., Bott, M., Griesinger, C. and Becker, S. (2008) A ligand-induced switch in the periplasmic domain of sensor histidine kinase CitA. *J Mol Biol* **377**(2): 512-523.

Shah, N., Gaupp, R., Moriyama, H., Eskridge, K.M., Moriyama, E.N. and Somerville, G.A. (2013) Reductive evolution and the loss of PDC/PAS domains from the genus Staphylococcus. *BMC Genomics* **14**: 524.

Shannon, C.E. (1949) Communication in the presence of noise. *Proc. Institute of Radio Engineers* **37**(1): 10-21.

Shaulsky, G., Fuller, D. and Loomis, W.F. (1998) A cAMP-phosphodiesterase controls PKA-dependent differentiation. *Development* **125**(4): 691-699.

Shimizu, M., Kato, H., Ogawa, T., Kurachi, A., Nakagawa, Y. and Kobayashi, H. (2010) Sigma factor phosphorylation in the photosynthetic control of photosystem stoichiometry. *Proc Natl Acad Sci U S A* **107**(23): 10760-10764.

Shoumskaya, M.A., Paithoonrangsarid, K., Kanesaki, Y., Los, D.A., Zinchenko, V.V., Tanticharoen, M., Suzuki, I. and Murata, N. (2005) Identical Hik-Rre systems

are involved in perception and transduction of salt signals and hyperosmotic signals but regulate the expression of individual genes to different extents in synechocystis. *J Biol Chem* **280**(22): 21531-21538.

Sidote, D.J., Barbieri, C.M., Wu, T. and Stock, A.M. (2008) Structure of the Staphylococcus aureus AgrA LytTR domain bound to DNA reveals a beta fold with an unusual mode of binding. *Structure* **16**(5): 727-735.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D. and Higgins, D.G. (2011) Fast, scalable generation of high‐quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **7**(1).

Smith, H. (2000) Phytochromes and light signal perception by plants--an emerging synthesis. *Nature* **407**(6804): 585-591.

Soderling, S.H. and Beavo, J.A. (2000) Regulation of cAMP and cGMP signaling: new phosphodiesterases and new functions. *Curr Opin Cell Biol* **12**(2): 174-179.

Sola, M., Gomis-Ruth, F.X., Serrano, L., Gonzalez, A. and Coll, M. (1999) Three-dimensional crystal structure of the transcription factor PhoB receiver domain. *J Mol Biol* **285**(2): 675-687.

Song, Y., Peisach, D., Pioszak, A.A., Xu, Z. and Ninfa, A.J. (2004) Crystal structure of the C-terminal domain of the two-component system transmitter protein nitrogen regulator II (NRII; NtrB), regulator of nitrogen assimilation in Escherichia coli. *Biochemistry* **43**(21): 6670-6678.

Sorzano, C.O., Marabini, R., Velazquez-Muriel, J., Bilbao-Castro, J.R., Scheres, S.H., Carazo, J.M. and Pascual-Montano, A. (2004) XMIPP: a new generation of an open-source image processing package for electron microscopy. *J Struct Biol* **148**(2): 194-204.

Stark, C., Breitkreutz, B.J., Reguly, T., Boucher, L., Breitkreutz, A. and Tyers, M. (2006) BioGRID: a general repository for interaction datasets. *Nucleic Acids Res* **34**(Database issue): D535-539.

Stepto, R.F.T. (2010) Dispersity in polymer science (IUPAC Recommendation 2009). *Polymer International* **59**(1): 23-24.

Stock, A., Koshland, D.E., Jr. and Stock, J. (1985) Homologies between the Salmonella typhimurium CheY protein and proteins involved in the regulation of chemotaxis, membrane protein synthesis, and sporulation. *Proc Natl Acad Sci U S A* **82**(23): 7989-7993.

Stock, A.M., Mottonen, J.M., Stock, J.B. and Schutt, C.E. (1989) Three-dimensional structure of CheY, the response regulator of bacterial chemotaxis. *Nature* **337**(6209): 745-749.

Stock, A.M., Robinson, V.L. and Goudreau, P.N. (2000) Two-component signal transduction. *Annu Rev Biochem* **69**: 183-215.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., Kuhn, M., Bork, P., Jensen, L.J. and von Mering, C. (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* **43**(Database issue): D447-452.

Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I. and Ludtke, S.J. (2007) EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol* **157**(1): 38-46.

Taylor, B.L. and Zhulin, I.B. (1999) PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiol Mol Biol Rev* **63**(2): 479-506.

Terashima, M., Petroutsos, D., Hudig, M., Tolstygina, I., Trompelt, K., Gabelein, P., Fufezan, C., Kudla, J., Weinl, S., Finazzi, G. and Hippler, M. (2012) Calcium-dependent regulation of cyclic photosynthetic electron transfer by a CAS, ANR1, and PGRL1 complex. *Proc Natl Acad Sci U S A* **109**(43): 17717-17722.

Thomason, P. and Kay, R. (2000) Eukaryotic signal transduction via histidine-aspartate phosphorelay. *J Cell Sci* **113 (Pt 18)**: 3141-3150.

Tomomori, C., Tanaka, T., Dutta, R., Park, H., Saha, S.K., Zhu, Y., Ishima, R., Liu, D., Tong, K.I., Kurokawa, H., Qian, H., Inouye, M. and Ikura, M. (1999) Solution

structure of the homodimeric core domain of Escherichia coli histidine kinase EnvZ. *Nat Struct Biol* **6**(8): 729-734.

Tullberg, A., Alexciev, K., Pfannschmidt, T. and Allen, J.F. (2000) Photosynthetic electron flow regulates transcription of the psaB gene in pea (Pisum sativum L.) chloroplasts through the redox state of the plastoquinone pool. *Plant Cell Physiol* **41**(9): 1045-1054.

Umena, Y., Kawakami, K., Shen, J.R. and Kamiya, N. (2011) Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9 A. *Nature* **473**(7345): 55-60.

Van Heel, M. (1987) Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction. *Ultramicroscopy* **21**(2): 111-123.

van Heel, M., Gowen, B., Matadeen, R., Orlova, E.V., Finn, R., Pape, T., Cohen, D., Stark, H., Schmidt, R., Schatz, M. and Patwardhan, A. (2000) Single-particle electron cryo-microscopy: towards atomic resolution. *Q Rev Biophys* **33**(4): 307-369.

van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R. and Schatz, M. (1996) A new generation of the IMAGIC image processing system. *J Struct Biol* **116**(1): 17-24.

van Heel, M., Keegstra, W., Schutter, W.G., van Bruggen, E.F.J. (1982) Arthropod hemocyanin studied by image analysis. *Life Chem Rep Suppl.* **1**: 69-73.

Velikova, N. and Wellsa, J.M. (2016) Rationale and Prospects of Targeting Bacterial Two-component Systems for Antibacterial Treatment of Cystic Fibrosis Patients. *Curr Drug Targets*.

Vidal, R., Lopez-Maury, L., Guerrero, M.G. and Florencio, F.J. (2009) Characterization of an alcohol dehydrogenase from the Cyanobacterium Synechocystis sp. strain PCC 6803 that responds to environmental stress conditions via the Hik34-Rre1 two-component system. *J Bacteriol* **191**(13): 4383-4391.

Voet-van-Vormizeele, J. and Groth, G. (2008) Ethylene controls autophosphorylation of the histidine kinase domain in ethylene receptor ETR1. *Mol Plant* **1**(2): 380-387.

Vuillet, L., Kojadinovic, M., Zappa, S., Jaubert, M., Adriano, J.M., Fardoux, J., Hannibal, L., Pignol, D., Vermeglio, A. and Giraud, E. (2007) Evolution of a bacteriophytochrome from light to redox sensor. *EMBO J* **26**(14): 3322-3331.

Wagner, J.R., Brunzelle, J.S., Forest, K.T. and Vierstra, R.D. (2005) A light-sensing knot revealed by the structure of the chromophore-binding domain of phytochrome. *Nature* **438**(7066): 325-331.

Wagner, J.R., Zhang, J., Brunzelle, J.S., Vierstra, R.D. and Forest, K.T. (2007) High resolution structure of Deinococcus bacteriophytochrome yields new insights into phytochrome architecture and evolution. *J Biol Chem* **282**(16): 12298-12309.

Wang, H., Yang, Y., Chen, W., Ding, L., Li, P., Zhao, X., Wang, X., Li, A. and Bao, Q. (2013) Identification of differentially expressed proteins of Arthrospira (Spirulina) plantensis-YZ under salt-stress conditions by proteomics and qRT-PCR analysis. *Proteome Sci* **11**(1): 6.

Wang, L.C., Morgan, L.K., Godakumbura, P., Kenney, L.J. and Anand, G.S. (2012) The inner membrane histidine kinase EnvZ senses osmolality via helix-coil transitions in the cytoplasm. *EMBO J* **31**(11): 2648-2659.

Wang, S. (2012) Bacterial Two-Component Systems: Structures and Signaling Mechanisms. Protein Phosphorylation in Human Health. C. Huang, *INTECH*.

West, A.H. and Stock, A.M. (2001) Histidine kinases and response regulator proteins in two-component signaling systems. *Trends Biochem Sci* **26**(6): 369-376.

Whitmore, L. and Wallace, B.A. (2004) The Peptaibol Database: a database for sequences and structures of naturally occurring peptaibols. *Nucleic Acids Res* **32**(Database issue): D593-594.

Whitmore, L. and Wallace, B.A. (2008) Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers* **89**(5): 392-400.

Willett, J.W. and Kirby, J.R. (2012) Genetic and biochemical dissection of a HisKA domain identifies residues required exclusively for kinase and phosphatase activities. *PLoS Genet* **8**(11): e1003084.

Winkler, K., Schultz, A. and Schultz, J.E. (2012) The S-helix determines the signal in a Tsr receptor/adenylyl cyclase reporter. *J Biol Chem* **287**(19): 15479-15488.

Wojnowska, M., Yan, J., Sivalingam, G.N., Cryar, A., Gor, J., Thalassinos, K. and Djordjevic, S. (2013) Autophosphorylation activity of a soluble hexameric histidine kinase correlates with the shift in protein conformational equilibrium. *Chem Biol* **20**(11): 1411-1420.

Wu, S. and Zhang, Y. (2007) LOMETS: a local meta-threading-server for protein structure prediction. *Nucleic Acids Res* **35**(10): 3375-3382.

Wurgler-Murphy, S.M. and Saito, H. (1997) Two-component signal transducers and MAPK cascades. *Trends Biochem Sci* **22**(5): 172-176.

Yoshida, T., Cai, S. and Inouye, M. (2002) Interaction of EnvZ, a sensory histidine kinase, with phosphorylated OmpR, the cognate response regulator. *Mol Microbiol* **46**(5): 1283-1294.

Zapata, L., Ding, J., Willing, E.M., Hartwig, B., Bezdan, D., Jiao, W.B., Patel, V., James, G.V., Koornneef, M., Ossowski, S. and Schneeberger, K. (2016) Chromosome-level assembly of Arabidopsis thaliana Ler reveals the extent of translocation and inversion polymorphisms. *Proceedings of the National Academy of Sciences of the United States of America* **113**(28): E4052-E4060.

Zapata, L., Ding, J., Willing, E.M., Hartwig, B., Bezdan, D., Jiao, W.B., Patel, V., Velikkakam James, G., Koornneef, M., Ossowski, S. and Schneeberger, K. (2016) Chromosome-level assembly of Arabidopsis thaliana Ler reveals the extent of translocation and inversion polymorphisms. *Proc Natl Acad Sci U S A* **113**(28): E4052-4060.

Zhang, Y. (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**: 40.

Zhao, F.Q. and Craig, R. (2003) Capturing time-resolved changes in molecular structure by negative staining. *J Struct Biol* **141**(1): 43-52.

Zhao, H., Piszczek, G. and Schuck, P. (2015) SEDPHAT--a platform for global ITC analysis and global multi-method analysis of molecular interactions. *Methods* **76**: 137-148.

Zhou, Y.F., Nan, B., Nan, J., Ma, Q., Panjikar, S., Liang, Y.H., Wang, Y. and Su, X.D. (2008) C4-dicarboxylates sensing mechanism revealed by the crystal structures of DctB sensor domain. *J Mol Biol* **383**(1): 49-61.